

Ethics and fraud

The trajectory of the Hwang scandal highlights the shortness of the path between unethical behaviour and outright misconduct.

The fall of Woo Suk Hwang represents perhaps the highest-profile case in the sorry history of research misconduct. The sheer Shakespearian drama of the Korean cell biologist's eclipse, surrounded by fawning courtiers and plotting groups of acolytes and enemies and in full view of the television cameras, has left few researchers of any discipline, anywhere in the world, unaware of its circumstances.

But what really makes the Hwang case special is the importance of the impugned results (W. S. Hwang *et al.* *Science* **303**, 1669–1674; 2004 and **308**, 1777–1783; 2005). The claimed cloning of human embryos placed Hwang at the forefront of stem-cell research, perhaps the most acclaimed and contentious sphere of modern science. It is also the first major misconduct case to occur during the modern era of carefully patented biology, in which scientists may aspire not just to fame, but to fortune as well.

All of this raises some general questions about ethics and fraud that researchers had perhaps hoped were put to bed a long time ago. Many of these questions were last publicly addressed in the long and painful aftermath of the Baltimore case, in which a junior researcher, Thereza Imanishi-Kari, was accused of fraud in the laboratory of one of the United States' leading microbiologists.

David Baltimore was eventually vindicated and is today president of the California Institute of Technology. But when the allegations were made, the National Institutes of Health opened an Office of Scientific Integrity, which was later downgraded but continues to support fraud investigations at US universities while seeking to get academics to teach their students about ethics and misconduct.

"The US system for dealing with fraud, imperfect as it may be, is still more advanced than that of many other nations."

This system, imperfect as it may be, is still more advanced than that of many other nations. Elsewhere in the world, cases of fraud have highlighted considerable shortcomings in the mechanisms for responding to misconduct.

Another celebrated fraud case involved Jan Hendrik Schön, a physicist at Bell Laboratories in Murray Hill, New Jersey, who was found guilty in 2002 of fabricating results in a massive and influential string of papers on superconductivity. However, the general view of the Schön case, outside the discipline directly involved at least, has been that no innocent civilians were hurt, and the fraud was unearthed and duly punished. There is almost a nagging sympathy for how such a smart young man could be so stupid.

The scope of the Hwang scandal tends to leave these and other recent misconduct cases in the shade, however. Although the full facts of the case remain unknown, its multidimensional outline is already clear. Barely a stone was left unturned in his lab's attempts to secure a positive result. Egg donors were energetically sought from

every conceivable source — including from junior researchers inside the laboratory itself, where allegations of coercion have been made. In the end, a result was claimed and the paper published in the complete absence of genuine scientific evidence to support it.

This will surely leave some people asking: if this single cell in the body of science was so malignant, how fares the rest? This is an awkward question and one that many in the community will seek to evade by referring to the fact that it happened in South Korea and couldn't happen here (wherever 'here' happens to be). And it is true that standards of oversight at many laboratories would, at least in theory, make this kind of scenario improbable. However, research — not least in the dynamic and fiercely competitive field of reproductive biology — is increasingly conducted through international collaborations. These often involve working with colleagues in countries, including China and South Korea, where mechanisms for supervising ethics and investigating misconduct are at relatively early stages of development.

"This will surely leave people asking: if this single cell in the body of science was so malignant, how fares the rest?"

It therefore falls to scientists who take research ethics seriously to confront the question head on, and to determine what steps should be taken to redouble our efforts to secure standards in both the ethics and conduct of research.

Hand in hand

A first step is to acknowledge that sound ethics and good research practice go hand in hand. The international stem-cell community showed some reluctance to distance itself from Hwang when serious questions were raised in this journal in May 2004 about the manner in which eggs had been obtained for his experiments (see *Nature* **429**, 3; 2004).

As soon as his main US collaborator, Gerald Schatten of the University of Pittsburgh, announced in November that he was bailing out of his collaboration with Hwang (see *Nature* **438**, 262–263; 2005), people began to speculate that Schatten must know there was a problem with the result of the seminal 2004 paper. After all, they inferred, no one would leave a wildly successful research group over ethical transgressions. Or would they?

The leadership of the scientific community has long argued for a solid line to be drawn between ethical transgressions — not informing patients of their rights, sexually harassing staff, coercing junior colleagues, that sort of thing — and actual research misconduct, which refers to the fabrication, falsification or plagiarism of scientific evidence.

In the wake of the Baltimore scandal, a congressionally mandated commission chaired by Kenneth Ryan called for a broader definition



of research misconduct that would embrace some forms of malfeasance beyond fabrication, falsification and plagiarism. His definition didn't cover bioethics, but it did class any breach of research regulations as misconduct.

Ryan's proposal was roundly condemned in the community, which fought a lengthy and successful battle to derail it. Researchers feared that the extension of misconduct investigations to embrace all kinds of professional laxity would lead to endless, fruitless investigation and, in particular, elicit groundless allegations from junior laboratory malcontents.

It is certainly true that there's a distinction between personal misbehaviour in the lab and outright scientific fraud, and it is perhaps

as well that special investigative procedures are retained exclusively for the investigation of the latter. Furthermore, the question of what constitutes an ethical transgression may vary between societies that elect to impose different rules, whereas scientific fraud knows no borders.

In view of the pattern of behaviour that led up to Hwang's disgrace, however, no one should argue ever again that despotism, abuse of junior colleagues, promiscuous authorship on scientific papers or undisclosed payment of research subjects can be tolerated on the grounds of eccentricity or genius. Research ethics matter immensely to the health of the scientific enterprise. Anyone who thinks differently should seek employment in another sphere. ■

Three cheers for peers

Thanks are due to researchers who act as referees, as editors resolve their often contradictory advice.

There is nothing like a high-profile fraud case to encourage journals to reflect on whether the standards and procedures they follow in selecting work for publication are thorough and appropriate. Misconduct creates a negative perception of journals' scientific peer-review processes, and the Hwang fraud saga has already fuelled some misconceptions about how the combination of referees and journal editors actually works.

Peer review remains by far the best available system for scientific quality control, however, and is an ultimately inspiring one at that. *Nature* is hugely grateful for the advice it receives from about 6,000 referees each year — typically two or three referees per paper. Most of their reports contain exactly what we need: a statement of what the referee considers the central message of the paper; an assessment of its significance; and a critique of technical or interpretational weaknesses, either in the work itself or in its presentation.

Between them, these elements add up to a verdict on the work's credibility and robustness. The system, it should be noted, is reliant on trust that what is written in the paper is actually true: it is not designed to detect the tiny minority of papers that are fraudulent.

The lives of editors and editorial boards of all journals are made interesting by the fact that, in many cases, the referees disagree on the verdict. And authors tend to be mightily upset if their papers are rejected when one or more reviewers are positive. Why, they demand to know, should the view of a negative reviewer be allowed to dominate the editors' selection decision?

To shed some light on how these decisions are reached, it is worth reflecting on some case studies of how and why referees differ in their view of papers submitted to *Nature* and the *Nature* research journals.

In one case, an exciting result relied on two techniques and a theoretical interpretation. The theoretical referee was very positive because the work validated an interesting idea. A specialist in one of the techniques was positive because he could find no flaw in its application. But the third referee uncovered a technical shortcoming in the second technique, and the paper was rejected after the editor assessed the significance of the shortcoming.

In another case, one of the referees recommended publication of

a paper, but also pointed out limitations in the value of the finding. The editor concluded that the paper lacked the significance that would justify inclusion in the journal.

On many other occasions, however, the editors' discretion in making a decision results in a paper's publication. In one such case, referees criticized a molecular-biology paper for a lack of mechanistic insight and expressed reservations about the appropriateness of some of the techniques the authors used. But the editors felt that the therapeutic implications of the paper merited publication and, after resolving the technical issues raised by the referees, pushed ahead with publication of what turned out to be a highly cited development.

Another paper concerned the innovative application of chemistry to an environmental problem. But before publication, the editor orchestrated considerable iteration between referees from quite disparate backgrounds to ensure that a common understanding of the paper and its reliability had been established.

In another case, efforts to obtain review of a paper in genetics led to seven refusals to review, one damning review and only one positive review. In this case, the editor identified an experiment that would improve the paper and suggested it, yielding interesting results that were then published and well received.

Only in a minority of cases does every referee agree on whether or not to publish a paper. The above examples illustrate just a few ways in which such differences arise and demonstrate why journals would lose the respect of their authors and readers if they were to act robotically on the referees' advice. Moreover, *Nature* often makes referees aware of what the others are recommending, which can sometimes provide useful feedback to the selection process.

We would never claim that *Nature's* decision-making process is perfect. Its imperfections — along with those of every other journal — are among the multifarious reasons why over-reliance on journal publications as a measure of researchers' performance is dangerous. *Nature* has also dished out its fair share of historically embarrassing rejections (see *Nature* 425, 645; 2003). We can only work to ensure that what we publish will do justice to the diversity of expertise that is brought to bear on its selection, and, above all, stand the test of time. ■

"Only in a minority of cases does every referee agree on whether or not to publish a paper."

RESEARCH HIGHLIGHTS

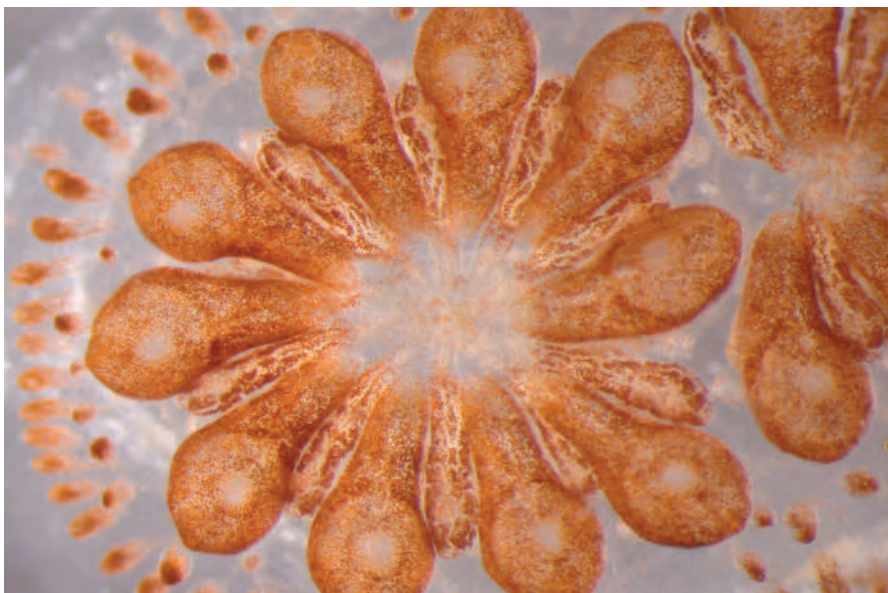
The ultimate parasite

Cell **123**, 1351–1360 (2005)

A cramped colony of sea squirts can use a single stem cell to take over the reproductive function of a neighbouring colony, researchers at Stanford University in California have found.

When two colonies of genetically dissimilar sea squirts (*Botryllus schlosseri*, pictured) come into contact, they repel each other. But compatible colonies fuse, with their vascular networks joining to form a single circulatory system.

Diana Laird and her colleagues discovered that, in fused systems, blood-borne stem cells from one colony can give rise in the other to gametes, the units of sexual reproduction, or polyps, for asexual reproduction. The results suggest that genetically 'fit' invading stem cells can outcompete native stem cells — showing how evolutionary pressures can come to bear on stem cells.



D. LAIRD

ARCHAEOLOGY

Written history

Science doi:10.1126/science.1121745 (2006)

The earliest-known example of writing from the Mayan civilization, which flourished in Central America from 1,000 years BC, has emerged from the ruins of a temple at San Bartolo in Guatemala.

A fragment of wall painted with a column of ten hieroglyphs (illustrated, right) was recovered from deep within a pyramid-shaped structure at the site. Radiocarbon dating of burnt wood found in portions of the structure built earlier, later and at the same time as the section where the writing was found put its age between 200 and 300 years BC, report William Saturno of the University of New Hampshire in Durham and his colleagues. The find suggests that the Maya evolved writing systems around the same time as other Meso-American civilizations.



MARINE ECOLOGY

Fishing prohibited

Science **311**, 98–101 (2006)

Counter-intuitively, no-fishing zones designed to save large Caribbean fish such as the Nassau grouper (*Epinephelus striatus*) also benefit threatened corals, reveals a survey of the

waters of the Bahamas archipelago.

Peter Mumby of the University of Exeter, UK, and his international team had been worried that more grouper would mean fewer of their prey — parrotfishes (Scaridae) — which graze on algae that otherwise smothers the coral. Fortunately, though, the fishing-free area also saved large species of parrotfish from the net. These big fish can escape the grouper and are naturally the busiest grazers. The

scuba-diving scientists found twice as much grazing and four times less algae inside the reserves. It's a happy ending for the coral, says Mumby.

NANOMATERIALS

A rather odd film

Phys. Rev. Lett. **95**, 266101 (2005)

In electronic devices shrunk to the nanoscale, changing the thickness of a metallic film by just one atom can have profound effects. Some of these effects can be fine-tuned, as illustrated by Tai-Chang Chiang and his colleagues from the University of Illinois at Urbana-Champaign in a neat set of experiments.

They studied how the electronic energy levels, and hence the overall stability, of a lead film on a silicon substrate varied with its thickness. Films made from an even number of atomic layers were normally more stable than films with an odd-numbered thickness. But the authors found that introducing a single layer of indium atoms

between the lead and silicon reversed the trend, so that odd-numbered layers were the more stable.

CANCER BIOLOGY

Bad influence

J. Clin. Invest. **116**, 261–270 (2006)

A chaperone protein that normally 'nannies' other proteins in times of stress could also contribute to breast cancer, say researchers in the United States. Vincent Cryns of Northwestern University in Chicago and his team studied a protein known as α B-crystallin, which stops damaged proteins from clumping together and helps others to refold correctly. It was present at high levels in cells from an aggressive type of breast cancer — the basal-like subtype.

The team thinks that the protein switches on a signalling system that involves the enzyme MAPK kinase. They showed that adding extra α B-crystallin to human breast cells made them develop abnormalities typical of the cancer, and that blocking the identified signalling pathway prevented these changes.

NEUROSCIENCE

Seeing is believing

Neuron **49**, 81–94 (2006)

Bipolar cells in the retina, which enhance the contrast of visual input, have curious properties. In one subpopulation, for example, the neurotransmitter GABA has opposite effects on the membrane charge at the cell's two ends. It depolarizes the input end while hyperpolarizing the end that

stretches towards the retina's output layer.

Thomas Euler and Thomas Kuner of the Max Planck Institute for Medical Research in Heidelberg, Germany, and their colleagues confirm the long-standing hypothesis that the opposing responses are enabled by a gradient in the concentration of chloride ions running down the cell's length. To show this, they engineered mice to express Clomeleon, a chloride-sensitive fluorescent protein.

MOLECULAR BIOLOGY

Copyeditor stops press

Nature Struct. Mol. Biol. doi:10.1038/nsmb1041 (2005)
A class of enzyme thought to be no more than the cell's back-room copyediting desk in fact has a key role in controlling gene activity, say researchers.

The enzymes, called ADARs, edit RNA by altering specific RNA bases. This editing changes the function of several mammalian genes, but mysteriously, its main target is the large number of RNAs that do not code for proteins. Now Kazuko Nishikura of the Wistar Institute in Philadelphia and her colleagues have an explanation.

They studied non-coding RNAs called microRNAs that switch off, or silence, target genes. They found that ADARs alter the processing of particular microRNAs and suppress their expression. This shows that RNA editing and RNA interference interact.

SMART MATERIALS

Paint that listens

J. Acoust. Soc. Am. **119**, 251–261 (2006)

It looks like paint, and can be sprayed like paint. But it's actually a microphone.

The mixture of ceramic nanoparticles,

polymer resin and light-emitting dye, concocted by James Gregory of Purdue University in West Lafayette, Indiana, and his co-workers, generates an image of a sound wave bouncing off a thin, porous film of the mixture applied to a surface. The brightness of the dye emission depends on the partial pressure of oxygen in the surrounding air, and so is a measure of air pressure.

Pressure-sensitive paints have been developed over the past several years for aeronautical studies, but have only recently been proposed as acoustic sensors. Gregory and his team show that their paint is sensitive enough to map the changing shape of a sound field over an area of about 22×17 centimetres.

ASTRONOMY

The constant question

Phys. Rev. Lett. **95**, 261301 (2005)

Astronomers are suggesting a way to tighten the noose around claims that a fundamental constant of physics has changed over the lifetime of the Universe.

Nissim Kanekar of the National Radio Astronomy Observatory in Socorro, New

Mexico, and his colleagues present an independent test of controversial data suggesting the fine-structure constant, α , has changed over the past 8.5 billion years. The team analysed absorption and emission lines from hydroxyl and hydrogen atoms in radio observations. They conclude that the fractional change in α over the past 6.5 billion years was less than 6.7×10^{-6} .

Although this new limit does not rule out the change in α claimed from observations of quasars, the data used are less prone to systematic errors.

A precise test of the original study's conclusions will be possible with the improved sensitivity of the next generation of radio telescopes.

PLANETARY SCIENCE

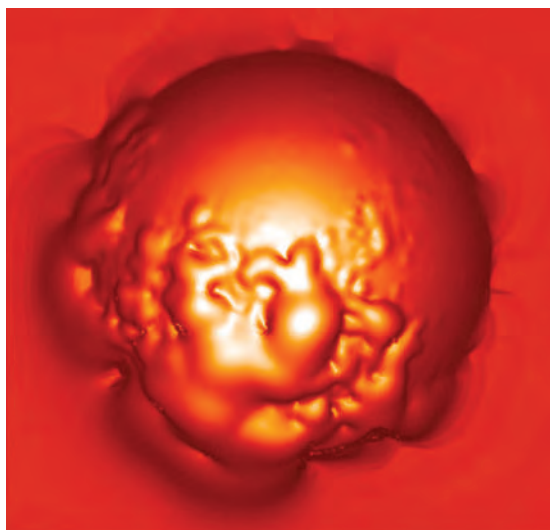
Martian lights

Geophys. Res. Lett. doi:10.1029/2005GL024782 (2006)

Aurorae regularly flicker on and off at Mars, a study suggests.

After the recent report of the first known martian aurora (*Nature* **435**, 790–794; 2005), David Brain of the University of California, Berkeley, and his colleagues delved into past data from the Mars Global Surveyor satellite. In six years' worth of data, they uncovered some 13,000 energetic electron-collision events linked to hundreds of auroral flashes.

The aurorae occur when electrons accelerate along magnetic field lines that connect the solar wind to localized magnetic fields in the martian crust (pictured above). Many of the most energetic aurorae took place during solar storms, suggesting that electrons are most readily accelerated during those times.



JOURNAL CLUB

Kurt Cuffey
University of California,
Berkeley, USA

A method for dating the carving of landscapes by ice or water captivates a glaciologist.

As an undergraduate, I authored a manuscript on the shapes of glacial valleys, and successfully had it rejected by two minor journals.

The silver lining of that cloud was the ensuing dialogue with wise geologists, who introduced me to

glacial geomorphology manuscripts founded in geophysics and rich with conceptual insight (B. Hallet *J. Glaciol.* **23**, 39–50; 1979; J. Oerlemans *Z. Gletsch. Glazial.* **20**, 107–126; 1984).

With my passion for the aesthetics of glacially sculpted terrain, it was thrilling to see great science applied to the subject. Yet equally striking was the weakness of the discipline's empirical side — a problem that persists.

Dramatic, characteristic landforms prove that glaciation has profoundly altered landscapes.

But how to quantify the magnitude and timing of the alteration?

David Shuster of the California Institute of Technology in Pasadena and his colleagues recently proposed a remarkable new source of information (D. Shuster *et al. Science* **310**, 1668–1670; 2005).

Erosion of a land surface cools the underlying rocks, in turn restricting the mobility of helium produced *in situ* by radioactive decay. Shuster *et al.* say they can reconstruct detailed cooling histories from the spatial

distribution of helium-isotope ratios within single mineral grains.

They apply their technique to samples from the walls of a glacial valley in British Columbia, Canada. It reveals that the rocks underwent rapid cooling 1.8 million years ago, suggesting the valley was deeply and rapidly scoured at that time.

If the method withstands scrutiny, studies of active processes, such as my own investigations of how glaciers erode the cores of mountain belts, will finally be complemented by precise dating of landform evolution.

NEWS

Verdict: Hwang's human stem cells were all fakes

SEOUL

The results are in. The university committee looking into scientific misconduct in the laboratory of South Korean cloner Woo Suk Hwang announced on 10 January that his 2004 claim to have cloned a human embryo was fake. But his Afghan hound Snuppy is a real clone.

The announcement finally confirms the gravest suspicions of Hwang's work with humans. There are two papers in which Hwang's group claimed to clone human cells — a 2004 article that describes the first cloned embryo and derivation of a stem-cell line from it (W. S. Hwang *et al. Science* **303**, 1669–1674; 2004), and a 2005 article that claims the establishment of eleven 'patient-specific' stem-cell lines (W. S. Hwang *et al. Science* **308**, 1777–1783; 2005). Both have turned out to be complete and deliberate fakes.

"Such an act is nothing other than deception of the scientific community and the public at large," concludes Myung Hee Chung of Seoul National University (SNU), who headed the committee.

With the 2005 paper already discredited in the panel's interim report (see *Nature* **439**, 8; 2005), Chung's statement focused on the 2004 paper. DNA fingerprinting tests carried out by three laboratories found that the genetic material of the supposedly cloned human cell line, NT-1, did not match that of the donor. Nor did it match any of the stem-cell lines from the *in vitro* fertilization (IVF) embryos of MizMedi Hospital, which were the source for the faked data in the 2005 paper.

Further investigation revealed that mitochondrial DNA from the cell line matched one of the egg donors, but the DNA inside the cells' nuclei varied at several locations. The committee concluded that the line was derived by parthenogenesis — where the single set of chromosomes in an egg develop as if it were fertilized. The images and data in the paper that showed perfect matches were fabricated.

The committee also found that Hwang worked with a staggering number of eggs — 2,061 from 129 women — despite claiming to

have used only 242 eggs for the 2004 study and 185 for the 2005 study.

The findings are a huge setback for therapeutic cloning — the idea that cloned embryos could be used as a source of patient-matched stem cells to replace damaged tissues in a range of diseases. Even using numbers of human eggs of which other researchers can only dream, Hwang's team was unable to derive such stem cells, and the field is now left with no evidence that it is possible in humans at all (see *Nature*, **438**, 1056–1059; 2005).

The committee did find that Hwang succeeded in cloning human embryos to the blastocyst stage, from which stem cells can be derived. But the success rate was just 10%, and they were "in poor condition". The only other group to have some success, Alison Murdoch's team at the University of Newcastle upon Tyne, UK, has cloned just a single blastocyst (M. Stojkovic *et al. Reprod. BioMed. Online* **11**, 226–231; 2005).

It is possible to create embryonic stem-cell



At least Snuppy has been confirmed as a clone.



lines, insists Kevin Eggan, a researcher in the field at Harvard University, Massachusetts. But no one will venture a guess as to when it might be accomplished. "There are many unknowns," says Eggan. "We don't know how many eggs will be needed and we don't know how many women will step forward to contribute."

Ethical transgressions in the way Hwang got his eggs — he seems to have coerced junior researchers into donating — have stim-

French research chief quits over reforms

PARIS

Bernard Meunier, president of France's basic-research agency, the CNRS, resigned on 5 January. The move brings to a head simmering internal tensions over the future of the 26,000-member body.

The CNRS has declined to comment on the resignation, apart from issuing a short statement by Meunier. In it, he makes public his disagreement over the reform plans of the agency's director-general, mathematician Bernard Larrourou. In principle, the president defines the general goals of CNRS policy and the

director-general carries them through, but in practice the latter holds the reins of power.

The reforms came into force on 1 January. They are meant to encourage multidisciplinary, wealth creation, the development of labs outside Paris, and tighter links with French universities. The structural reforms are due to be completed later this year by a 'strategic plan' that will lay out future policies in more detail.

As a result of the reform, the CNRS's eight existing departments have been regrouped into four broad departments — life sciences,

chemistry, humanities, and maths and physics — and two 'cross-cutting' departments — engineering and the environment, and sustainable development.

Meunier, a chemist, regards Larrourou's reforms as unnecessary management interference that he believes will weaken science at the agency. He thinks that the new configuration of departments would complicate rather than simplify matters, with laboratories often belonging to several different departments at once, and he questions how the cross-

**SPECIAL REPORT: HWANG**

All our news on the scandal, complete with a timeline and a guide to who's who.

www.nature.com/news/specials/hwang



Myung Hee Chung announces that claims in two *Science* papers were deliberately fabricated.

ulated an international debate over how eggs should be obtained. Egan expects to gain approval this spring to begin human stem-cell cloning research, and he says his group will follow the US National Academies' guidelines. These stipulate that egg donors should receive no payment.

Even when admitting faked data, Hwang has maintained that his human cloning techniques are valid. But most experts say they merely involve tweaks to previously known methods, such as squeezing the nucleus out of cells rather than sucking them out with a needle. "Besides some slight adjustments, there was really nothing new," says Teruhiko Wakayama of the Center for Developmental Biology in Kobe, Japan, who created the first mouse clones in 1998.

Many experts conclude that Hwang's greatest achievement was Snuppy, the first cloned dog (B. C. Lee *et al.* *Nature* **436**, 641; 2005). The SNU investigators verified Snuppy's identity as a clone by proving that he had the same nuclear DNA as the skin-cell donor and the same mitochondrial DNA as the egg donor — a conclusion that was confirmed on 10 January by *Nature's* own investigation.

Dog ovulation produces very immature eggs, so culturing them is difficult, even for basic IVF, says Wakayama. "If it's real, this is their greatest accomplishment," he says. The SNU committee also noted that Hwang — originally trained as a veterinarian — showed greatest skill when it came to cloning animals, notably pigs and cows.

As *Nature* went to press, however, the Munhwa Broadcasting Corporation, which originally accused Hwang of faking his data, was about to air a television programme questioning Hwang's claim to have cloned a cow — the work that first shot him to fame in South Korea. The SNU committee said it was unable

"Hwang merely tweaked previously known methods. There was really nothing new."

to confirm whether the cow was a clone because Hwang did not cooperate with them.

The committee, which issued a 50-page report covering the investigation, stopped short of accusing Hwang or other individuals on the team of deliberate fabrication. It deferred to national prosecutors who will now look into legal aspects of the case and the possibility of fraud. Hwang received huge funding from the Korean government for his work, including an annual stipend of US\$3 million, which he started receiving this year as the country's "supreme scientist".

The Chung committee did report, however, that Hwang was lying when he said in November that he did not know about egg donations by his lab members. "Hwang accompanied the student to the hospital himself," the report says. Hwang later circulated a form asking for voluntary egg donation and collected signatures from female technicians.

While the committee delivered its report, a small group of supporters at the university entrance held up signs demanding that Hwang be given a chance to prove himself. "We'll put the smile back on your face," read the slogan on a familiar billboard calling Hwang the "Pride of Korea".

Chung does not agree. Hwang's team, he says, "cannot represent science in Korea". But he ends on a hopeful, if defensive, note. "We are certain that this learning experience will be a stepping stone for better execution and management of scientific research."

David Cyranoski

D. CYRANOSKI

cutting departments would work.

The reforms also create five regional CNRS boards, and Meunier argues that this would add an unnecessary layer of bureaucracy, and hand excessive power to the regions, weakening scientific imperatives from central management. "He feels that it risks creating five little CNRSs," says Jacques Fossey, a chemist who is head of the research union SNCS and a member of the CNRS board of directors.

In his resignation statement, Meunier slams the reforms as "an excessively administrative network" and not the "simple and dynamic mode of functioning" that the



Bernard Meunier has resigned over changes to France's research agency.

agency needs. He adds that he hopes his resignation will lead to a new team that is "more adapted to the actions the CNRS needs to take to assure its place in French and European research". Fossey

believes Meunier intended to provoke a crisis, gambling that this would force the government to remove Larrouturou and appoint a new management team.

The upset comes just as the research agency's dominant role in French science is in question. The CNRS funds its own labs, but research will increasingly be driven by competitive grant proposals administered by the National Research Agency, a body set up last year with an initial budget of €350 million (US\$423 million), which will rise to €1.5 billion by 2010.

The CNRS's power would be further reduced by a long-awaited research and innovation reform bill,

due to be voted on by the National Assembly next month. This would create a national Agency for Research Evaluation, responsible for looking at all research agencies, labs and scientists, a job currently done by the CNRS national committee.

As *Nature* went to press, the French government was tipped to appoint as Meunier's successor Catherine Bréchnignac, a physicist and president-elect of the Paris-based International Council for Science. Bréchnignac was director-general of the CNRS from 1997 to 2000, and earned a reputation as a staunch defender of the agency's autonomy.

Declan Butler

C. LEBEDINSKY/CNRS PHOTO THÉQUE

Yes, but will it jump?

As the world braces itself for a potential flu pandemic, experts are wrestling with the key question: whether the H5N1 virus is likely to evolve the ability to spread between people. Several researchers are arguing that the risk has been overplayed. But others are now attacking this view, stating that there is no evidence to support such a claim.

H5N1 has led to the death or slaughter of hundreds of millions of poultry in Asia, and is spreading relentlessly among birds worldwide. It has killed about half of the 156 people so far identified as having been infected, and last week it claimed its first human victims outside Asia — at least two youngsters in a family in eastern Turkey (see 'Bird flu takes its toll on Turkey').

Despite the apparent threat, some virologists, including Peter Palese of the Mount Sinai School of Medicine in New York and Paul Offit of the University of Pennsylvania School of Medicine, believe that much of the attention focused on H5N1 is unwarranted. They argue that H5 viruses may be inherently incapable of transmitting efficiently from human to human. The viruses have had ample opportunity to mutate into a pandemic strain, they argue, and if they haven't already done so they probably never will.

Of the 16 H subtypes of flu, only H1, H2 and H3 are known to have caused human pandemics, including the most recent ones in 1918, 1957 and 1968. "There is a possibility that another subtype may jump among people," Palese says. "However, that has never been observed."

But other experts have told *Nature* that

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Clusters of bird flu in Turkey have increased fears of possible human-to-human transmission.

this cannot be used to predict the chance of H5N1 or any future H5 strain triggering a pandemic, and that to state otherwise sows unnecessary confusion.

So who is right? Much of Palese and Offit's argument stems from the idea that many asymptomatic cases of H5N1 have gone undetected. The reported death rate of about 50% is "a complete exaggeration, and not scientifically justified," says Palese.

"If there were indeed many other people infected without showing symptoms then Peter might have a point," says Albert Osterhaus, a

virologist at the Erasmus Medical Center in Rotterdam, the Netherlands. But he says that is far from proved. Studies to find out how many people in the general population of affected countries have antibodies to H5N1 have only just started. Osterhaus says that the first, so far unpublished, results suggest that large numbers of poultry workers do not seem to have been infected. "The data we have suggest that the clinical cases are not the tip of the iceberg, but rather the only people who have been infected," he says.

Palese argues that other evidence suggests

M. OZER/AFP/GETTY

Bird flu takes its toll on Turkey

At least two people have died from avian flu in Turkey. And, as *Nature* went to press, the country's health ministry had confirmed that a further 12 people are infected with the H5N1 virus. Together, these are the first reported human cases outside Asia. Avian flu first arrived in Europe's birds last July, when outbreaks occurred in Russia. It has since appeared in birds in Kazakhstan,

Croatia, Romania and Turkey.

The death of a 14-year old boy from Dogubayazit, a town on Turkey's border with Iran, on 1 January, was followed by that of his 15-year-old sister on 5 January. Their 12-year-old sister died the following day, although lab tests have yet to confirm H5N1 as the cause of her death. The children's six-year-old brother is currently in hospital.

H5N1 infection has since been confirmed in two other children from another family living in the region, and some 30 suspected cases in the area and in the northern Black Sea region are under surveillance. Last weekend, the Turkish health ministry announced three cases in Ankara, 1,000 km west of the first outbreak. Two brothers, aged five and three, and an unrelated 65-year-old

man are thought to be infected.

Like many of the recent outbreaks in Asia, the Turkey cases have occurred in family clusters, raising the worrying prospect of limited human-to-human transmission. This is extremely difficult to prove, as it requires working out how the first case got infected, then ruling out the possibility that other family members got ill the same way.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

H5 strains have been infecting humans from birds for decades. In particular, he cites a paper from 1992, which found H5 antibodies in 2–7% of the population in rural China (K. F. Shortridge *Semin. Respir. Infect.* 7, 11–25; 1992). There is no reason then to believe that an H5 virus should suddenly trigger a pandemic now, he says.

Edward Holmes, who studies virus evolution at Pennsylvania State University, is unconvinced by Palese's argument. "I don't think you can say that it should have happened already," he says. "The history of evolution is that rare things can happen." And in unpublished work, Yi Guan of the University of Hong Kong, who

worked on the 1992 study, says that his group has since tested 4,000 blood samples collected in southern China in 2001 and found no antibodies against H5 viruses.

Experimental evidence about whether H5N1 could spread between people is also lacking. Although scientists are beginning to understand what genes make flu viruses pathogenic, they know little about what determines their transmissibility.

Where a virus is excreted from seems to be one factor. Viruses infecting the upper respiratory tract, for example, are thought to jump more easily than those lower down. But scientists aren't sure about this with H5N1, as only a handful of autopsies have been done on human victims, largely because of religious concerns in the relevant countries.

Another way to study the question would be to study viruses such as the 1918 flu strain in mammals, replacing the genes one at a time to investigate their roles in transmissibility. This work hasn't been done either, because such dangerous viruses are strictly regulated.

More data are desperately needed, says Holmes, and in the meantime we should not assume that H5N1 won't cause a pandemic. "Trying to put numbers on how long it is going to take or how many people are going to die is a pointless exercise. We scientists should just hold up our hands and say we don't know what is going to happen." Palese himself accepts this view as reasonable. "OK; that I can agree with," he says.

Palese also strongly supports the need for better surveillance and research. Whatever he thinks about H5N1, he emphasizes that further flu pandemics are inevitable. His views have been picked up by the media as evidence that the pandemic threat is being hyped, but Palese argues that the need to prepare has not been driven home strongly enough: "People are just not willing to accept that we are not doing enough across the board — in research, surveillance or in developing vaccines." ■

Declan Butler

It's hard to say how significant the clusters are, says Albert Osterhaus, a virologist at the Erasmus Medical Center in Rotterdam, the Netherlands. Families are likely to be exposed to the same things, he points out, "so you wouldn't necessarily expect a random distribution of cases".

Scientists are also intrigued by the fact that in the clusters, it tends to be blood relatives

who get infected, rather than, say, husband and wife. This suggests that some people might be genetically more susceptible than others. But there are too few cases to confirm this, says Peter Palese, a virologist at the Mount Sinai School of Medicine in New York. It could just depend on who prepares the food, or kills the chickens.

He adds that research on

genetic differences among victims could be used to look for differences in, for example, the immune response — although researchers are rarely able to get hold of such data.

But the new cases in Europe do not fundamentally increase the chances of a pandemic, says Osterhaus. "With 142 cases in Asia, a couple of new ones on the doorstep of Europe doesn't change much," he says. **D.B.**

ON THE RECORD

“This is the National Hurricane Center signing off for 2005...finally.”

A forecaster at the hurricane centre in Miami, Florida, breathes a sigh of relief that Tropical Storm Zeta has finally dissipated, more than a month after the hurricane season ended.

“Like the Loch Ness monster, this can bring some good for our country.”

A tourist official in Malaysia hopes for a flood of visitors now that Bigfoot has apparently been seen there.

Source: National Hurricane Center, Associated Press

SCORECARD

**Asthma**

For once, your mobile phone could be good for your health. Some London-based networks are offering a text-messaging service that alerts asthma sufferers to bad-air days.

**Being called Li**

A survey of 300 million Chinese people has revealed that the country's most popular surname is Li, a name held by 7.4% of the population. Second and third ranked are Wang and Zhang, held by 7.2% and 6.8% of people, respectively.

**Booze in space**

A rumour that Russia will bring alcohol aboard the International Space Station is quashed by teetotal NASA.

NUMBER CRUNCH

A report by Germany's Centre for University Development has led newspaper *Die Zeit* to suggest that German universities could maybe learn a trick or two from their US counterparts when it comes to raising additional funds.

\$540 million was raised by Harvard University in 2003 — the most raised by any US university that year.

\$16 million was raised by the University of Mannheim — the most raised in Germany in 2003.

50% of German universities do not do any fundraising at all.

DNA tests put death penalty under fire

WASHINGTON DC

US campaigners against capital punishment are hoping that modern DNA tests in old cases will undermine public confidence in the death penalty.

In the next few weeks, genetic testing of an old sample could show that Roger Keith Coleman — executed in 1992 for rape and murder — was not guilty. The tests are being conducted at Canada's Centre of Forensic Sciences in Toronto, and were requested by the state of Virginia's governor, Mark Warner. If Coleman is exonerated, it will be the first time that genetic evidence is used to prove that the US death penalty has killed an innocent person.

Coleman was convicted in Virginia for the 1981 murder of his sister-in-law. The sample being tested was gathered in 1990 as part of an unsuccessful attempt to clear his name. In ordering the recent tests, Warner — a Democrat widely expected to run for president in 2008 — called Coleman's case “an extraordinarily unique circumstance, where technology has advanced significantly and can be applied in the case of someone who consistently maintained his innocence until execution”.

Only once before in the United States has DNA been tested in a case where the prisoner had already been executed. The tests, performed in 2000, were inconclusive about the guilt of the Georgia man, Ellis Felker, who had been executed four years earlier.

But modern DNA tests have successfully exonerated some Virginia prisoners. Between 2002 and 2004, three people were cleared after tests were done on old biological samples, saved by a methodical employee in the era before such techniques were available. As a result, Warner asked for tests on a random 10% of all the samples that had been saved. Out of thousands of cases, there were 29 in which current DNA tests could have shown guilt or innocence. Two of the 29 were shown to be innocent and their convictions overturned.

Warner has now announced that the rest of the samples will also be tested, calling this “the only morally acceptable course”. Ellen Qualls, the governor's communications director, says

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

that a rough extrapolation from the test results suggests there may be 20 or 30 exonerations from around 300 pertinent cases — an error rate that some would argue is unacceptable for a state that allows capital punishment.

The US public has long been divided over the death penalty. Each state decides whether to administer it or not: 38 states allow people to be sentenced to death, and Illinois has suspended executions while it examines the issue. Proof that innocents have died would certainly strengthen arguments against the penalty.

The issue of wrongful convictions does not apply just to cases that pre-date DNA testing. A growing scandal in Texas serves as a reminder that genetic tests are only as reliable as those carrying them out. On 4 January, a damning

report was published on the Houston Police Department's crime laboratory. The DNA unit of the Texas lab was closed in 2002, after a series of exposés by local television news station KHOU. Now, an independent investigation, which has looked at 1,100 cases so far, charges the lab with problems including contamination of samples, misrepresentation of

“Retesting DNA samples is the only morally acceptable course.”



CONFERENCE REPORT

See our website for news from the American Astronomical Society meeting.

www.nature.com/news

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

TEK IMAGE/SPL

Innocents at risk: the United States is in a furore over questionable DNA evidence.

statistics in court and failure to use controls. In two cases, one of which put a man on death row, a clear absence of the suspect's genetic markers in a sample was called "inconclusive" and not reported.

Investigators blamed lack of funding and inept management, among other factors. "The complete lack of outside scrutiny of the crime lab's operations, procedures, and reporting of results allowed serious deficiencies...to become so egregious that analysts in the lab simply had no perspective on how bad their practices were," the report reads. DNA analysed at the lab is being retested, and at least one person has been freed.

Nearly 95% of US police crime laboratories are accredited by a board that maintains national standards. Ralph Keaton, head of the accreditation project, says that Houston's experience is the exception rather than the rule. "From the beginning," he says, "DNA evidence has been carefully scrutinized because it is so powerful."

Emma Maris

Bird lovers keep sharp eye on owls

Should wild animals be reintroduced into areas where they have become extinct? In recent years, the issue has split European conservationists and farmers, who disagree over the future of species such as wolves. Now the unexpected arrival of a tiny population of owls in Britain has divided even advocates of conservation.

The breeding pair of eagle owls (*Bubo bubo*) first attracted attention last autumn, when they featured in a television documentary. Many ornithologists believe that the birds, which have a wingspan of two metres, haven't lived in Britain since the last ice age, 10,000 years ago. The owls are fearsome hunters, happy to dine on mammals — sometimes as large as small deer — or on other birds of prey.

The Royal Society for the Protection of Birds is calling on the British government to monitor the new arrivals, resident in Yorkshire in northern England, in case the population grows. The society is worried that the owls might start to kill significant numbers of endangered species, such as black grouse and hen harriers.

But some researchers argue that eagle owls should be protected as native birds. John Stewart, a palaeontologist at University College London, says fossil evidence suggests that the owls survived previous ice ages. He argues that remains from the

past 10,000 years should be taken as evidence that the owls existed in the wild — ornithologists had previously assumed that the remains came from tame eagle owls that had been imported for hunting. "If you look at the



Soar point: should the eagle owl be allowed to repopulate Britain?

owls' distribution around Europe, it is like that of other animals such as wolves and beavers that have been pushed out of northwest Europe," adds Stewart. "It's churlish to suggest that they don't belong here."

If he is right, eagle owls could be added to the British List, the register of birds recorded in recent times in Britain and Ireland. This would give them the right to be protected, and perhaps even actively reintroduced. But Tim Melling, secretary of the records committee at the British Ornithologists' Union, which compiles the list, says he is not convinced by Stewart's points.

Melling says he and colleagues have reviewed every published British bird book and found no evidence for an established population of eagle owls. He points out that the birds have been popular pets for hundreds of years and that reports of wild eagle owls often note that the animals are unusually tame. A similar debate over belonging was ignited last year by carbon dating of bones of the Eurasian lynx (*Lynx lynx*)

from caves in Yorkshire. The lynx was thought to have died out in Britain several thousand years ago, when climate change turned much of its forest habitat into boggy peatlands. But the bones were just 1,500 years old, suggesting that recent human activity was responsible for the animals' demise.

The result was a boost for conservationists wanting to bring the lynx back to Britain, including David Hetherington, who led the carbon-dating study at the University of Aberdeen. He suggests that forestry projects and deer in the Scottish Highlands would provide suitable habitat and prey.

But whatever the fossil evidence, any reintroduction needs local backing, cautions Toby Aykroyd, vice-chairman of the Wilderness Foundation, a charity based in Chelmsford, Essex. Farmers in France have protested against the return of wolves, which prey on livestock. And beavers, which have been successfully reintroduced in many European countries, have proved controversial in Britain: farmers fear they will be not be allowed to remove the animals from their land.

Jim Giles

D. NILL/NATURE PL

Methane finding baffles scientists

MUNICH

The startling discovery that terrestrial plants produce the greenhouse gas methane is sending scientists in two disciplines, not to mention a few politicians, back to the drawing board.

The newly revealed methane emissions have taken plant physiologists by surprise, because far more energy is required to create methane than, say, carbon dioxide in an oxygenated environment. Climate researchers are also amazed that they could have missed what is potentially a huge methane source — up to a third of all methane produced worldwide (see 'How could we have missed this?').

Until now, it was thought that plant matter produces methane only through microbial activity in oxygen-free environments such as swamps, flooded rice fields and ruminants' guts. But on page 187 of this issue, Frank Keppler, a geochemist at the Max Planck Institute for Nuclear Physics in Heidelberg, Germany, and his colleagues report that grasses and leaves from various species release the gas under normal aerobic conditions.

The source of the methane — and why plants make it — is unknown. Some species make other volatile hydrocarbons such as isoprene, but that reaction involves a specific enzyme, and only seems to kick in when the plants need to dissipate excess energy. The methane emissions that Keppler found rise smoothly with temperature up to 70 °C, suggesting that no enzyme is involved.

"This seems to be a secondary chemical reaction with no specific function for



Methane machine? Vegetation may be a huge source of this major greenhouse gas.

plant metabolism," says Elmar Weiler, a plant physiologist at Ruhr University in Bochum, Germany. "It's a truly surprising finding."

But beyond its implications for botany, the discovery could prove important for understanding and predicting climate change — and for our attempts to reduce greenhouse-gas emissions. Methane is the second most important greenhouse gas in the atmosphere after carbon dioxide, and levels have doubled over the past 200 years, mainly as a result of increased agricultural activity.

The finding doesn't change ideas about the total amount of methane being released into

the atmosphere. But scientists had thought they knew about all the significant methane sources and how much each contributed (see page 148). Now it seems that their figures were very wrong. As a rough estimate, Keppler reckons that global vegetation may be releasing between 60 million and 240 million tonnes of methane each year — up to a third of the total amount that enters the atmosphere.

"The surprising thing to me is the amount of methane they found," says Martin Heimann, director of the Max Planck Institute for Biogeochemistry in Jena, Germany. "It means we neglected a big driving force for the climate."

It is too early to say exactly how the revelation might influence predictions for future climate change, but it's unlikely to be good news. The fact that plant methane emissions rise with temperature, and that plants are likely to grow faster in a warmer climate anyway, could lead to a big rise in methane emissions from natural sources, says Johannes Lelieveld, an atmospheric researcher at the Max Planck Institute for Chemistry in Mainz, Germany.

The finding also restricts our options for reducing methane emissions, he points out, because measures such as growing rice in drier fields are likely to prove less effective than had been thought. "If natural greenhouse-gas sources are greater than we thought, the scope for climate politics becomes narrower," he says. "You wouldn't cut down forests just because trees release methane."

Quirin Schiermeier

Additional reporting by Mark Peplow.

How could we have missed this?

The finding that plants are a major source of methane has left many scientists struggling to believe it could have been missed before — and wondering what else might have been overlooked if it is true.

Keppler points out that detecting the methane was far from easy, as the amount released by individual plants is tiny compared with levels already in the atmosphere. His team were studying chemical reactions in ageing plants, and acted on a hunch after they found hints of methane from leaves left in an incubator. To

check their finding, they carried out studies in methane-free air, and irradiated plants to rule out microbial activity as a cause.

But the discovery has made climate researchers wonder how much they really understand about greenhouse-gas sources and sinks. "I don't think there will be many more big surprises," says Drew Shindell, a climate researcher at NASA's Goddard Institute for Space Studies in New York. "But I also wouldn't bet that this is the last one."

Others want to wait until it becomes clear exactly how the

methane is produced before they jump to any conclusions. "My feeling is that this could be very important," says David Beerling, a palaeoclimatologist at the University of Sheffield, UK. "But inferring a methane source by incubating leaves or placing chambers over plants can mean the nature of the source is quite uncertain."

"I don't know what to make of it," adds Colin Prentice, a biogeochemist at the University of Bristol, UK. "My first reaction is scepticism. I find it hard to believe that we missed this." **Q.S. & M.P.**

Budget cut hits core US biomedical research

Scientists funded by the US National Institutes of Health (NIH) may have something other than alcohol to blame for their New Year's headache: their first budget cut in 36 years.

On 30 December, President George W. Bush finally signed the NIH's 2006 budget. The agency will get \$28.6 billion in funding, a \$35-million reduction from last year. This cut comes despite the government agreeing an increase of \$206 million for the year, close to the amount requested by the President last February (see *Nature* 433, 559; 2005). But that increase was erased by a later government bill for a 1% across-the-board cut in spending on all discretionary programmes (that is, in spending that is not being channelled into existing programmes), in part to help pay for the war in Iraq.

Biology research advocates are lobbying for better funding in 2007, but with the US budget crunch continuing for the foreseeable future, prospects are bleak.

Phoenician ports lurk under modern cities

The easing of political tension in Lebanon has allowed archaeologists to locate the sites of ancient Phoenician ports.

The cities of Tyre and Sidon on the Lebanese coast dominated Mediterranean trade thousands of years ago and still exist today. But although the cities' names have not changed, the coastline on which they sit has been reshaped by silting over the past three millennia.

Nick Marriner of the European Centre for Research and Teaching on the Geosciences of the Environment in Aix-en-Provence, France, and his colleagues drilled a total of 40 sediment cores in the two cities. They report in the January issue of *Geology* that the old harbours lie underneath today's urban centres. Marriner hopes that his findings will help efforts to protect the cultural heritage of the two cities.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Cores brought up from Sidon city have unearthed clues about the old harbour that still lies beneath.

Physicist's snowflake images land on US stamps

The next Christmas holiday season in the United States will bring a small flurry of stamps featuring science. In October, the US postal service will issue a set of four stamps showing snowflake images taken by Kenneth Libbrecht, a physicist at the California Institute of Technology in Pasadena.

Libbrecht, who by day works on the LIGO gravitational-wave project, has perfected a cold-weather camera rig for photographing snowflakes quickly, before they melt.

The four images were taken in chilly weather in Alaska, Michigan and Ontario. Libbrecht's home page, www.snowcrystals.com, features more shots of both real and synthetic snowflakes.



2006 USPS

US bill outlaws political interference in science

The long-awaited budget bill that includes funding for the US National Institutes of Health (NIH) also contains a law banning political interference in scientific decisions.

Senator Richard Durbin (Democrat, Illinois) inserted the law in the annual appropriations bill for the Department of Health and Human Services (HHS), which includes the NIH and the Public Health Service. In 2004, the National Academy of Sciences found that candidates to federal scientific advisory panels were often subjected to political litmus tests.

The new law makes it illegal for nominees to HHS advisory panels to be questioned about their political affiliation or voting history, or for HHS officials to disseminate false or misleading scientific information. The measure, however, includes no penalties if broken.

Missouri team uncover record-breaking prime

Mathematicians have unveiled the largest prime number found to date, a 9.1-million-digit monster that, if printed, would fill about eight issues of *Nature*.

"There's almost no practical use for it," says Curtis Cooper, leader of the Central Missouri State University team that found it. "I just love the beauty of big primes."

Cooper's team is part of the Great Internet Mersenne Prime Search project, which discovered the previous four largest primes. This project uses the idle time of hundreds of Internet-linked computers to test all Mersenne numbers (two to the power of a prime, p , minus one, or $2^p - 1$) to

check if they are primes. The newest recordbreaker is $2^{30,402,457} - 1$.

One incentive for searchers is that whoever finds the first 10-million-digit prime number will win US\$100,000 from the San Francisco-based non-profit Electronic Frontier Foundation, whose aim is to promote cooperative computing.

Journal bans authors with secret ties to companies

A US medical journal has adopted a policy of temporarily blacklisting authors who fail to disclose conflicts of interest.

The *Journal of Thoracic and Cardiovascular Surgery* decided to act in December after learning that four of the authors on two papers published in the second half of last year did not declare their financial links to a company whose heart-surgery technology they had evaluated. The authors gave favourable reports of devices manufactured by AtriCure of Cincinnati, Ohio.

The journal plans to publish a correction and, in the future, will refuse to publish submissions from authors who are found to have withheld such conflicts. The ban will probably last one to two years, says Andrew Wechsler, the journal's editor and a surgeon at Drexel University College of Medicine in Philadelphia, Pennsylvania.

Correction

The Sidelines column in the 10 November 2005 issue (*Nature* 438, 136; 2005) incorrectly characterized a survey of intellectual-property issues by the American Association for the Advancement of Science (AAAS). The item should have said that 40% of survey participants, not 40% of all AAAS members, reported that their research was affected by difficulties in obtaining patented technology.

BASE INVADERS

Could viruses have invented DNA as a way to sneak into cells? **John Whitfield** investigates.

What with the threat of bird flu, the reality of HIV, and the general unseemliness of having one's cells pressed into labour on behalf of something alien and microscopic, it is small wonder that people don't much like viruses. But it's possible that we may actually have something to thank the little parasites for. They may have been the first creatures to find a use for DNA, a discovery that set life on the road to its current rich complexity.

The origin of the double helix is a more complicated issue than it might at first seem. DNA's ubiquity — all cells use it to store their genomes — suggests it has been around since the earliest days of life, but when exactly did the double spiral of bases first appear? Some think it was after cells and proteins had been around for a while. Others say DNA showed up before cell membranes had even been invented¹. The fact that different sorts of cell make and copy the molecule in very different ways has led others to suggest that the charms of the double helix might have been discovered more than once². And all these ideas have drawbacks. "To my knowledge, up to now there has been no convincing story of how DNA originated," says evolutionary biologist Patrick Forterre of the University of Paris-Sud, Orsay.

Forterre claims to have a solution. Viruses, he thinks, invented DNA as a way around the defences of the cells they infected³. Little more than packets of genetic material, viruses are notoriously adept at avoiding detection, as influenza's annual self-reinvention attests. Forterre argues that viruses were up to similar tricks when life was young, and that DNA was one of their innovations.

To some researchers the idea is an appealing

way to fill in a chunk of the DNA puzzle. Furthermore, the hypothesis should be testable through genomic studies and even lab experiments. But whether or not it pans out, Forterre's idea reflects an emerging consensus that viruses' diversity, mobility and capacity for rapid change has made them major players in some of the most important moments in life's evolution.

Small world

Most researchers think that before there was DNA, life stored its information in RNA, the double helix's more versatile chemical cousin. RNA is a slender, flexible molecule, usually made as a single strand. RNA molecules can contort in ways that allow them to catalyse chemical reactions, including in some cases their own replication. It is possible to imagine an 'RNA world' where the molecule does almost everything — catalysing reactions and replications that would today be catalysed by

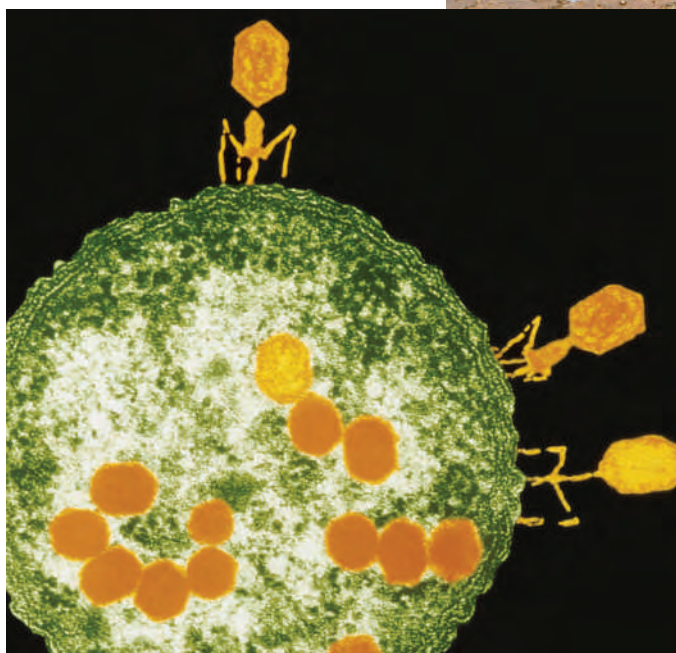
Hot springs give a glimpse of the wild diversity of viruses. But did these tiny experimenters stumble across DNA in their bid to invade cells?

proteins, and storing information that would now be stored in DNA. It is not possible to imagine DNA, a rigid, double-stranded rod that can be replicated only with the help of a protein, operating as a one-molecule band in the same way. That's one reason for thinking it came along later, after

complex proteins had been added to the RNA world.

Once DNA did arise, it would have had several advantages as genetic material. DNA's skeleton is more chemically stable than RNA. This stability allows DNA genomes to be longer, and so allows organisms to become more complex. But, as Forterre points out, the beneficial chemical property cannot explain why DNA appeared in the first place. Natural selection has no foresight; no innovator could acquire DNA on the basis that it would later be helpful in the expansion of its genome. "That would be like saying that dinosaurs evolved feathers because they knew they were going to turn into birds," says Forterre.

Instead, he thinks that DNA's original selective edge was that it allowed viruses to avoid their host's defences. Many cells repel invaders by degrading their genetic material. But enzymes that had evolved to break down RNA would have ignored DNA.





"A virus that invented DNA would have a tremendous advantage in overcoming cellular defences," agrees Malcolm White, a proteomics researcher at the University of St Andrews, UK. Forterre's hypothesis "fits with a lot of the clues that are scattered through genomes and phylogenies", he adds.

Insider traders

Two main lines of evidence point to viruses as likely genetic innovators, says Forterre. One is the diversity of genetic systems in contemporary viruses, which suggests an evolutionary tendency toward reinvention. Viruses have genomes made from double- and single-stranded DNA, double- and single-stranded RNA, and even DNA in which the chemical base uracil — also used in RNA — replaces the usual thymine. The genome can be carried on a single string, on a closed loop, or as a set of fragments. Many of these changes are thought to have occurred to help viruses avoid their host's defences.

In fact, biologists believe they are only just beginning to fathom the extent of this diversity. "We don't know much about the viruses of the world," says Philip Bell, a molecular biologist at Macquarie University in Sydney, Australia, who has argued that the nuclei found in complex cells are also descended from viruses⁴. Many of the viruses now being found in hot springs, for instance, feature unusual shapes — including spindles, rods, filaments and droplets — as well as genes with no known counterparts in other organisms⁵.

The other line of evidence rests on relations

among the genes used for DNA processing. There are three domains of cellular life: archaea, bacteria, and eukaryotes (the group containing plants and animals). All three share similar genes for turning genetic information into protein, and a similar enzyme for converting the components of RNA into those of DNA. This strongly suggests that these genes arose before the domains went their separate ways, probably in the RNA world.

But the similarities break down for DNA helicases and polymerases — the enzymes that unwind the DNA double helix and copy each strand. Although the archaeal and eukaryotic versions of these genes are similar, the bacterial versions are radically different from both, suggesting that perhaps these DNA replication systems evolved independently. What's more, the DNA polymerases of eukaryotes and bacteria are more closely related to similar enzymes found in viruses than they are to each other. This all implies to Forterre that the ability to copy DNA molecules did not originate with cells, but with their parasites.

In Forterre's scenario, the RNA genes in a cell infected with a DNA virus migrated to the new, more stable format over time with the help of a reverse transcriptase — an enzyme that makes DNA copies of RNA genes. Viruses could be expected to contain such enzymes, as they are so helpful for replication or for pinching useful genes from a host. Once the DNA genome became a complete warehouse of cellular genes, the original RNA chromosomes would be redundant. The process happened more than once, which explains the different DNA handling systems.

Anthony Poole, an evolutionary biologist at Stockholm University, Sweden, finds the idea intriguing. The hypothesis fits with much of the evidence from viral biology and gene trees, he says. "It's very well thought out. I like it a lot."

At the same time, Poole warns, the evidence is ambiguous. For instance, although viruses and cells swap genes with alacrity, it can be difficult to work out which genome gave birth to an innovation, and which imported it. "Phylo-

genies suggest a relationship between viral and cellular sequences. The problem is we don't know the order things happened in — viruses could derive from cellular lineages," he says.

Nor is everyone persuaded by Forterre's idea. Bill Martin, an evolutionary biologist at the University of Düsseldorf in Germany, says he agrees that the evidence points to DNA arising more than once, and that reverse transcriptase was probably involved in the transition from RNA to DNA. But he doubts that DNA's original selective advantage lay in infection. "It's completely off-target," says Martin. "The simple chemical stability of DNA is the main point."

Viral marketing

Forterre advocates gathering gene sequences from a greater diversity of viruses to seek descendants of the lineage that might have first infected cells with DNA. A good place to look, he says, would be a recently discovered group that infects amoebae, the mimiviruses⁶. These have huge, double-stranded DNA genomes — longer than those of some bacteria — and some of their genetic enzymes are similar to those in eukaryotes. Forterre also thinks that lab experiments with viruses and bacteria might recreate some aspects of the evolutionary process. It might be possible, for example, to replace a cell's DNA replication enzymes with their viral counterparts.

But can we ever really be sure about anything that happened so close to the origin of life? "It's an area of discourse rich in conjecture and poor in proofs, but I tend to be optimistic," says

"Up to now there has been no convincing story of how DNA originated."

— Patrick Forterre

Eugene Koonin. A genomics researcher at the National Center for Biotechnology Information, based in Bethesda, Maryland, Koonin was the first proponent of the idea that DNA evolved twice. He says the idea

of an RNA world shows that consensus — if not certainty — is possible, he says, and we can hope for the same regarding the origin of DNA. "The study of viral genomics is not going to come to fruition in the next five years, but it's not hopeless."

John Whitfield is a freelance science writer based in London.

1. Koonin, E. V. & Martin, M. *Trends Genet.* **12**, 647–654 (2005).
2. Leipe, D. D., Aravind, L. & Koonin, E. V. *Nucl. Acids Res.* **27**, 3389–3401 (1999).
3. Forterre, P. *Biochimie* **87**, 793–803 (2005).
4. Bell, P. J. L. *J. Mol. Evol.* **53**, 251–256 (2001).
5. Prangishvili, D. & Garrett, R. A. *Biochem. Soc. Trans.* **32**, 204–208 (2004).
6. Scola, B. *et al. Science* **299**, 2033 (2003).



T. CRADDOCK/SPL

L.D. SIMON/SPL

P. FORTERRE



Asteroid Itokawa dwarfs the shadow of a Japanese space probe (centre), whose mission was to collect a rock sample.

A shot in the dark?

Japan's mission to collect a sample from a distant asteroid looks to have ended in failure. **Ichiko Fuyuno** investigates how the setback will affect Japan's struggling space programme.

It was always a high-risk mission. No spacecraft has safely brought back a sample from the Solar System since the Soviet probe returned with lunar soil in the 1970s. So when, on 25 November 2005, a team from the Japanese space agency monitored the descent of the Hayabusa spacecraft towards the bumpy surface of the asteroid Itokawa, everyone in the control room was tense. Once Hayabusa was 360 metres above the asteroid, the touchdown command was issued. "I felt as if all the people in the room were riding on it and descending together," recalls Junya Tera-zono, the agency's publications officer, who was busy posting photos and live updates to a website as the spacecraft descended.

Despite the risks, after it had travelled 2 billion kilometres, and spent three months imaging the 540-metre-long rock, hopes that Hayabusa would bring back a souvenir from its trip were high. And on the morning of 26 November, a signal from the craft suggesting that it had fired pellets, designed to throw up rock fragments from the asteroid's surface,

caused an eruption of noise in Hayabusa's control room.

But the joy didn't last long. Just days later, the Japanese space agency, known as the Japan Aerospace Exploration Agency or JAXA, announced that it was highly unlikely that any pellets had been released or any sample collected. Mechanical problems had been detected in the probe back in July, but these troubles became catastrophic soon after the spacecraft landed on Itokawa. After the team lost communication with the spacecraft in early December, project manager Jun'ichiro Kawaguchi decided to delay Hayabusa's return by three years to 2010 to give them more time to revive it. The chances of a safe return look gloomy.

Hayabusa would have capped a break-

"If you want to climb Mount Everest or a small mountain, either way you have to move up step by step." — Masakazu Iguchi

through year for Japan's space programme — had everything gone well. After a difficult decade, marked by a string of expensive satellite and rocket failures and a tough budget environment, Japan merged its existing space agencies in October 2003.

The three agencies were the National Space Development Agency (NASDA) — Japan's main rocket and satellite developer; the Institute of Space and Astronautical Science (ISAS), responsible for scientific missions; and the smaller National Aerospace Laboratory. The merger was intended to cut costs and revitalize a space programme that had lost its way after a strong start in the 1970s and 1980s.

Lost in space

Today, the two-year-old JAXA has an ambitious wishlist for exploration over the next two decades, and a 2% budget increase for 2006 — the first budget increase for Japan's space programme in many years. But turning round Japan's fortunes in space exploration will depend on whether it can find ways to improve its track-record without killing its ambitious spirit.

Critics say Japan tries to do too much with too little. JAXA's budget (¥180 billion for 2006) is a tenth of NASA's, and less than half that of the European Space Agency or ESA (see graph opposite). And, at ¥12-billion (US\$100 million), Hayabusa cost only about half that of NASA's Stardust mission, which is set to return

AP/JAXA

SOURCE: EUROCONSULT

to Earth with captured cometary dust on 15 January. Japan can afford fewer missions, and so has fewer opportunities to launch new technologies. The result is to stuff as many ideas as possible into one launch. Hayabusa certainly carried a lot of hardware 'firsts'. Some of these, such as the Japanese ion-drive engine used to propel the spacecraft out to the asteroid, worked fine. Others, such as the small surface probe Minerva, failed to deliver.

Tales of woe

"Maybe sometimes Japan tries to do too much for its resources," says Andrew Cheng, a planetary scientist at Johns Hopkins University in Baltimore, Maryland, and a member of Hayabusa's science team. "I'm happy to see very brave decisions and to launch very complicated missions. All that is good," adds Cheng. "But they cannot fail every time either."

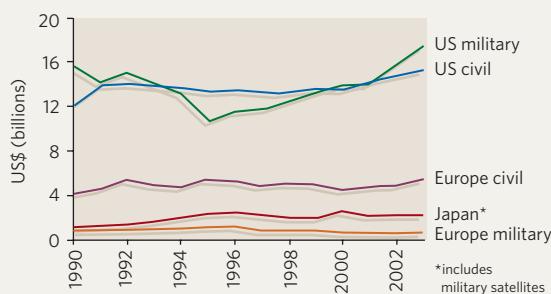
The year Hayabusa was launched was a particularly troubled time for Japan's space programme. In October 2003, the Midori-II Earth observation satellite failed. The following month, one of the Japanese flagship rockets, an H-IIA, had to be destroyed in mid-flight. Then the Mars probe Nozomi, in trouble since 1998, was finally lost in December. And last summer, the main X-ray instrument on the joint US-Japan Suzaku telescope shut down, reducing scientists' ability to study black holes.

Despite these troubles, many Japanese space experts believe that Japan should not just try to catch up with Europe and the United States, but should blaze its own trail. "Having ambitious dreams is good," says Masakazu Iguchi, head of the space activities commission that reviews Japan's space activities for the education ministry, which oversees JAXA's budget. But, he warns, "Japan should move steadily towards its goals. If you want to climb Mount Everest or a small mountain, either way you have to move up step by step." Iguchi argues that the important thing is to learn from failure. "I think JAXA understands that," he says.

Under pressure to improve the performance of Japan's space programme after the 2003 disasters, politicians sought the help of outside experts, including top US and European space-agency chiefs. And despite resistance from JAXA officials, the agency formed an advisory commission for mission success in 2004. Headed by former NASA chief Daniel Goldin, the commission released a report in March 2005, listing 21 ways the agency could improve.

The Goldin commission suggested that JAXA strengthen ties with industry by shifting technical responsibilities to its prime manufacturers. In the past, Japan's space programme retained

A DECADE OF SPACE BUDGETS



control over most design decisions, and interactions between agencies and the manufacturers were limited. It is hoped that with more responsibility, Japanese firms will gain the expertise needed to allow the country to compete in the global satellite market.

Another key recommendation was to boost the efforts of systems engineers. Toshifumi Mukai, who heads a chief engineer's office established in October 2005, says systems engineers do important work at the start of a project by defining mission requirements and identifying potential risks. Under the new system, chief engineers operate independently of the project managers, who are now required to share development data with others more openly.

But some JAXA officials are concerned that too much focus on risks, as well as constant reviews, will further weaken morale. "Just how to get prepared in the event of failures is becoming daily work. I think that's wrong," says Kawaguchi, who believes Japan must keep being adventurous. Many Japanese space experts are wary of adopting the approach taken by China's space programme. Although China has had two successful astronaut missions, it uses off-the-shelf technology, which many Japanese space experts dismiss as lacking innovation.

Others worry that JAXA will become as cautious as NASA or ESA. "I think ESA is more conservative than JAXA, at least as reflected in design philosophy for spacecraft and in mission operations," Cheng says. He hopes Japan does not become too risk averse.



False hope: Jun'ichiro Kawaguchi (centre) and his team eagerly await signals from their spacecraft as it descends to an asteroid.

There is no sign of that in JAXA's 20-year vision for space exploration, released in April last year. Calling for lunar exploration and perhaps eventually manned spaceflight, the ambitious scope of the 20-year plan seems at odds with current funding levels. Since a peak in 1999, the Japanese space budget has shrunk by 20%.

Risk taker

Critics, including the Goldin commission, have long argued for a strategic vision for Japan's space programme — one that will help it set priorities, and that will encourage better integration of the agencies that make up JAXA.

Since the merger, the three agencies have largely retained their separate cultures and resisted being unified further. The vision document is a first important step, says John Logsdon, professor of space policy at George Washington University in Washington DC. "JAXA is right now going through the process to deal with bureaucratic reorganization," he says. "It takes time."

Decisions about human spaceflight and Moon bases won't be made anytime soon, so JAXA can focus on immediate priorities, such as improving rocket reliability, says Kimikazu Iwase, director of the space development and utilization division at the education ministry. Iwase attributes a successful H-IIA rocket launch in February 2005, the first for 15 months, to better pre-launch testing.

Whatever Hayabusa does next, Kawaguchi's team has many busy months ahead analysing the data and images sent back by the craft before its descent. More than 1,500 high-resolution pictures have revealed a rocky surface devoid of debris. This is in striking contrast to the highly weathered surface of the asteroid Eros, which NASA's Shoemaker spacecraft visited in 2001.

Hayabusa did not achieve everything JAXA hoped for, but few question its engineering and scientific achievements. "Whether or not we ultimately get a sample returned to Earth, the mission still is a success from a science point-of-view," says Donald Yeomans of NASA's Jet Propulsion Laboratory in Pasadena, California, and US project scientist for Hayabusa. "The Japanese flight team performed well dealing with unexpected spacecraft anomalies and a bizarre and rocky asteroid surface."

What JAXA learns from such experiences will shape its fortunes over the next decade. "Overall, things are getting better, but we haven't fully gotten out of the doldrums," says Yasunori Matogawa, associate executive director at JAXA. "Hayabusa was the mission that could have opened the door. Now we will have to see whether it has really done so."

Ichiko Fuyuno is a contributing correspondent based in Tokyo.

KYODO

THE SHAPE OF THINGS TO COME

A number of fatal brain diseases are linked to misfolded proteins, an effect researchers are mimicking in the lab. But as they generate new versions of these malformed molecules, could they be creating a monster? **Roxanne Khamsi** finds out.



In a secure lab in Texas, five machines are purring away quietly. Working through the night, these boxes churn out billions of malformed proteins.

A seemingly odd thing to mass-produce, these distorted molecules are at the heart of research into a family of diseases that destroy the brain.

The equipment was devised by Claudio Soto¹, a biologist at the University of Texas Medical Branch in Galveston, who works on prion protein, a naturally occurring molecule. Misfolded versions of this protein are thought to cause conditions such as mad cow disease and its human equivalent, Creutzfeldt–Jakob disease (CJD). Such abnormal prions are infectious and, when ingested, make their way to the brain, where they slowly distort the natural prion proteins, causing disease and, ultimately, death.

But this conversion process takes a long time — in humans it can be more than a decade before enough abnormal prion has been made to cause disease, making it hard for scientists to investigate the diseases. Soto's machines have changed that: they can quickly turn a mass of normal proteins into twisted imitations of their former selves.

"We can mimic the process of prion replication that normally takes a year in the brain within a matter of hours in the test tube," says Soto. And that acceleration is allowing researchers to experiment with the abnormal prions, gaining fresh insight into how they cause disease. But it begs one important question: is it safe?

Collectively, the diseases linked to infectious prions are called transmissible spongiform encephalopathies, or TSEs, because of the spongy appearance they give the brain. Most TSEs are species-specific. Scrapie, for example, occurs in sheep; CJD in humans; and chronic wasting disease (CWD) in deer and elk. But some TSEs have hopped from one species to another — mad cow disease, for example, is believed to have jumped the 'species barrier' to cause disease in both cats and humans.

Given that the abnormal prions are notoriously hard to destroy — they resist temperatures as high as 120 °C and are not broken down by the enzymes that usually degrade proteins — is it wise to be creating new types of them in the lab?

Like all prion research, Soto's work is carried out under strict biosafety conditions. He argues that the risk of any prions escaping is very low. "The potential

benefits are much greater than the risk," he argues, adding that the work his team is doing could help predict whether a TSE is likely to jump between species.

Other researchers agree. They say that making prions in the test tube will help them tackle key problems such as why one type of abnormal prion behaves differently from another. "It's one of the most burning questions in the entire field," says Adriano Aguzzi, a prion researcher at the University Hospital of Zurich in Switzerland.

Scientists have long explored the question of whether TSEs can jump between species. Although a particular TSE spreads efficiently within a species, researchers have found it much harder to infect one species with the abnormal prion from another. This led to the idea of a species barrier that limited such infections in the wild.

Broken barriers

But that view was challenged in the late 1980s, when Britain's cow herds were struck by a new TSE: mad cow disease, or bovine spongiform encephalopathy. Scientists suspect that the species jump was caused by the farm animals being given feed created from the rendered carcasses of scrapie-infected sheep. And the disease didn't stop there — it also seemed to jump from cows to humans, causing a condition called variant CJD.

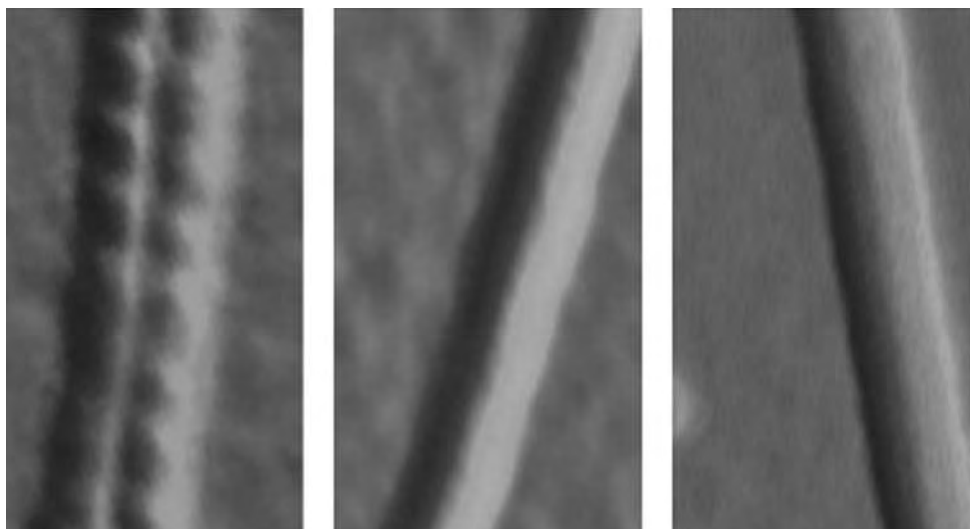
Experts generally agree that abnormal prions from cows can convert normal ones in humans, and cite this as an example of a prion disease crossing the species barrier. So just how easily do infectious prions jump from one animal species to another? It's no idle question. Although governments have banned certain animal by-products from feed to restrict the transmission of prion disease among livestock, some experts say that there may be a need for further restrictions to prevent the possible spread of TSEs to species such as pigs, which are currently thought to be much less susceptible to prion disease under experimental conditions.

But in the 1990s, a new TSE emerged in UK cats, claiming the lives of more than 80 domestic animals. It even spread as far as zoo animals, reaching cheetahs, ocelots, pumas, lions and tigers. Again it was linked to the BSE prion, suggesting that the species barrier might be easier to breach than was once thought.

To get a clearer picture of infectious prions and their ability to jump species, researchers are turning to Soto's mass-production technique. In the brain, the abnormal prions tend to link together in chains, eventually forming visible tangles. But this limits the number of prions

Rapid response: Claudio Soto demonstrates his equipment for mass-producing abnormal prion proteins.

REF. 7



Twisted into shape: just minor changes to a normal human prion protein (left) can change its three-dimensional structure (right) to resemble the infectious, disease causing prion found in hamsters (centre).

free to convert normal proteins — only those at the end of the chains can cause the switch, making it a very slow process. In his protein-misfolding cyclic amplification (PMCA), Soto found a neat way to sidestep this problem.

The technique initially incubates a small amount of abnormal prion with an excess of normal protein, so that some conversion takes place. The growing chain of misfolded protein is then blasted with ultrasound, breaking it down into smaller chains and so rapidly increasing the amount of abnormal protein available to cause conversions. By repeating the cycle, the mass of normal protein is rapidly changed into misfolded prion.

Tangled web

Soto and his team have used the technique to probe the limits of the species barrier. In unpublished work, they incubated normal protein from lab animals such as mice with a range of infectious prions, including those that cause mad cow disease, scrapie and CWD. The resulting prions caused disease when injected into the brains of rodents. “The data from our experiments suggest that there is no absolute species barrier. You can always break it,” says Soto. “It’s just a question of how far you push the system.”

Soto’s lab is now extending these experiments to look at other possible species jumps, including that from infected deer to humans. Although eating meat from scrapie-infected sheep does not seem to cause prion disease in humans, the effects of eating venison contaminated with CWD remain unclear. It has not yet been proved that CWD can jump to humans, but studies have shown that injecting CWD prions into monkeys can cause disease², says Paul Brown, former medical director of

“The concept of an absolute species barrier is probably a myth.” — Paul Brown

the Laboratory of Central Nervous System Studies in Bethesda, Maryland. He stresses that CWD does not pose a public health risk as deer remains do not typically get cycled into the food chain.

“The concept of an absolute species barrier is probably a myth,” says Brown, adding that there are simply “varying degrees of ease” in terms of how readily the prions can be transmitted. Under experimental conditions, skunks, pigs and even gerbils seem susceptible to TSEs from other species³. As yet, researchers have not succeeded in infecting dogs or rabbits with a TSE.

Taking the strain

Another big question is what makes a particular prion ‘strain’ act in the way it does. Strains are defined by the characteristics of the illness they cause in a given animal, including incubation time and the type of damage done to the brain. Often, different strains will have distinctive biochemical properties — although two strains within the same species can have the same amino-acid sequence and behave differently.

Byron Caughey and his team at the Rocky Mountain Laboratories in Hamilton, Montana, have been working on making prions in the test tube for more than a decade. In 1995, they showed that

infectious prions made in the absence of cells retain strain-specific qualities, such as tending to convert only prions from the same species^{4,5}, which suggests that no other cellular components are needed for the prion’s activity.

Some studies suggest that the species barrier is dictated largely by the three-dimensional structure of the abnormal prion species⁶. In other words, the given prion’s structure has to be able to work with the native normal protein. By changing just two key amino acids in the proteins, Witold Surewicz of Case Western Reserve University in Cleveland, Ohio, and his team have created new prion fragments that defy the species barrier^{7,8}. The result was new structures that could adopt the shape of prion fragments from a different species.

Using Soto’s technique, such investigations into the species barrier will only accelerate. But as biologists mix normal and infectious prions from different species in the lab, they will create novel prions with unpredictable abilities. The risks of creating a ‘super-strain’ are not new — scientists recently debated the merits of recreating and tinkering with the virus strain that caused the 1918 flu pandemic.

Proponents of the prion work say that the threat of creating an accidental monster is minimal. They stress that unlike the genetic manipulation of flu viruses, which could produce an inhaled superbug, similar work in prions is unlikely to produce infectious material that can be transmitted through the air.

Others point out that scientists have been creating new prion strains for many years — albeit in animals, rather than in the test tube. “I do not see the generation of new strains in cell-free systems as a fundamentally new type of biohazard,” says Caughey.

And understanding how easily particular infectious prions can pass from one species to another will help officials design adequate safety measures to protect us and our food chain, says Soto. “It’s very important to evaluate the danger and be prepared.” ■

Roxanne Khamsi reports for Nature from New York.

Pumas are among a few big cats in captivity to have contracted an illness similar to mad cow disease.



1. Saborio, G. P., Permanne, B. & Soto, C. *Nature* **411**, 810–813 (2001).
2. Marsh, R. F., Kincaid, A. E., Bessen, R. A. & Bartz, J. C. *J. Virol.* **79**, 13794–13796 (2005).
3. Brown, P. *Transmissible Spongiform Encephalopathy as a Zoonotic Disease* (International Life Sciences Institute Europe, Brussels, 2003).
4. Kocisko, D. A. et al. *Proc. Natl Acad. Sci. USA* **92**, 3923–3927 (1995).
5. Bessen, R. A. et al. *Nature* **375**, 698–700 (1995).
6. Peretz, D. et al. *Neuron* **34**, 921–932 (2002).
7. Jones, E. M. & Surewicz, W. K. *Cell* **121**, 63–72 (2005).
8. Vanik, D. L., Surewicz, K. A. & Surewicz, W. K. *Mol. Cell* **14**, 139–145 (2004).

BUSINESS

All systems go

Industrial chemists are borrowing techniques from drug researchers to track down materials with desirable properties. **Andrea Chipman reports.**

High energy prices and environmental worries are accelerating the hunt for good catalysts. Luckily, industrial chemists have found an unexpected tool to aid their search for better materials to speed up chemical reactions.

The latest addition to their repertoire comes from the drug industry and is known as combinatorial chemistry. The technique allows chemists to automatically synthesize and screen large numbers of compounds at once, helping to track down the ones that will prove useful. It has caught on fast in the petrochemical business, and industries making everything from cosmetics to food look set to exploit it too.

The technology has advanced significantly over the past five years, making it easier for industries to screen candidate materials automatically. The key to success is ensuring that small samples of a material will be reproduced accurately when production is scaled up.

"Assuming you have methods that are scalable, the speed comes from parallelization and automation," says Jason King, a chemist and head of business development at HTE, a company based in Heidelberg, Germany, that specializes in this high-throughput technology. Used the right way, the technique can bring down the cost of materials research and achieve in days or weeks what formerly took months or years.

"There's a huge increase in interest in advancing this technology," says King. "It has come of age at the right time."

High prices and profit margins in the energy business, together with growing demand for environmentally friendly processes, have encouraged major oil companies to buy into the new technology in a big way. They have signed large, long-term contracts with research companies such as Symyx Technologies of Santa Clara, California, which pioneered the approach.

Symyx was founded in 1994 by two biochemists: Alejandro Zaffaroni and Peter Schultz. The two scientists launched the company, which now employs 350 people, to find

new applications for high-throughput methods that were being used in pharmaceutical and genetic research. Zaffaroni had helped to set up Affymetrix, which supplies high-throughput genetic tools, and Schultz is director of the Novartis Research Foundation's genomics institute in San Diego.

Catalysts developed with high-throughput tools have found widespread use in the petrochemical industry, assisting the reactions that break down crude oil and gas, and convert the resultant hydrocarbons into plastics and other industrial products.

Much of the impetus for these advances comes from the fact that large corporations are looking strategically at all their research and development, and trying to raise productivity. "Clients are thinking about research and development at a different timescale and pace," says Isy Goldwasser, a chemical engineer and president of Symyx. "It takes a psychological shift."

Cleaning up

The nascent business of selling combinatorial chemistry techniques to the process industries is dominated by just three companies: Symyx, HTE and Avantium. Based near Amsterdam,

Avantium is the youngest, having been spun off from Shell Chemicals in 2000. It is developing high-throughput techniques to find suitable catalysts for processes such as hydro-treating — which uses hydrogen to remove contaminants such as sulphur from petrochemicals. Avantium says it has invested "tens of millions of euros" in developing kit that can screen up to 96 candidate catalyst materials at once.

HTE, which stands for high-throughput experimentation, got started a year earlier and is staffed by chemists from both industrial and academic backgrounds. At first, it sought to develop materials of its own, but now draws most of its business from selling software and equipment to larger companies and forming research collaborations with them. Last year, for example, the privately held firm signed a major collaboration of



Pioneering: Isy Goldwasser

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

undisclosed value with Albemarle, a Virginia-based supplier of refinery chemicals.

"There's interest from refineries because of legislative pressure to improve fuel specifications," explains King. "That's driving a lot of catalyst developers to look at taking this up."

The German company, which employs about 40 researchers, produces systems that can do a quick screen of up to 625 catalyst samples at once, or more thorough screens of up to 48 candidates. Both types of equipment use fixed-bed reactors, in which candidates are loaded into a pipe and gases or liquids passed through them to produce the desired reaction. The quick screens take as little as 20 minutes; in later phases, as the tests get more extensive, reactions can take weeks or even months.

HTE is also interested in using high-throughput techniques to develop cosmetics and other end products. It has devised equipment, such as a dispensing system for viscous liquids, to do this. The equipment can be used to produce as many as 96 different formulations a day, each of which is made up of different mixes of 10 to 15 liquid or powder components, King says. Because the process is completely automated, the results are more accurate and controllable than experiments done by hand.

King says that clients are keen to invest money to improve their research and development process. "Their real source of competitive advantage is technological innovation," he says. "The emphasis on technology is one of the reasons why the whole high-throughput

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

GETTY IMAGES

Pillars of the community: the petrochemical industry is benefiting from high-throughput techniques, and others will follow its lead.

area has seen such an upswing."

Being first off the blocks, Symyx has developed a wide range of patented materials, usually in partnership with clients. These include a phosphorus-based material that helps store X-ray images — developed with German film company Agfa — and a plastic used by Japanese semiconductor firm JSR. In total, it has struck deals with major industrial corporations that are worth more than US\$600 million, including a \$200-million, five-year research collaboration with ExxonMobil, the world's largest oil company, and a similar, \$120-million deal with Dow Chemical.

"Chemical corporations aren't used to working with small companies on research in the way pharmaceutical companies are," Goldwasser says. "Over ten years, we've developed a business model to find ways clients can work with us to solve tough problems. High-throughput technologies are applied to that problem, and the customer owns and commercializes the material while we get royalties based on the value we add."

Symyx's existing clients spend more than \$10 billion between them on research and development, leaving plenty of scope for expansion. "In my opinion," says Goldwasser, "we've only scratched the surface of ways high-throughput research can improve the effectiveness of research and development."

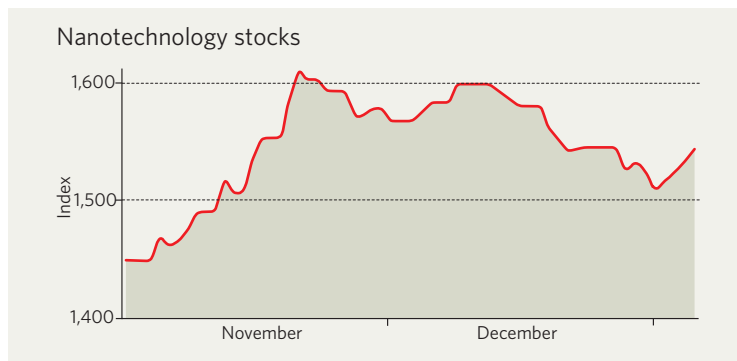
IN BRIEF

LEG UP Wyeth reached an agreement with Seattle biotechnology company Trubion to back its development of drugs based on molecules known as small modular immunopharmaceuticals, or SMIPs. The New Jersey drug company will invest \$40 million in Trubion immediately, with up to \$800 million to follow if certain milestones are met. Wyeth will gain most worldwide rights for marketing drugs developed under the deal, including a treatment for rheumatoid arthritis that is already in phase II trials.

INTEL OUTSIDE Semiconductor manufacturers announced plans to move their products — and their brand images — into people's living rooms. At a huge consumer electronics show in Las Vegas, Intel and its largest rival, AMD, said that their next generation of chips would serve as the brains of multimedia home entertainment systems. The companies will brand computers containing the chips, and seek to make them compatible with a range of televisions and stereos. Intel's chip is called Viiv (rhymes with 'five') and AMD's offering is called AMD Live.

HIGH POINT A British pharmaceutical company has won permission from the US Food and Drug Administration to conduct phase III trials of a cannabis-based drug for the treatment of cancer pain. The drug, Sativex, which is obtained from cannabis plants grown at an undisclosed location in Britain, has already been approved in Canada for the treatment of pain caused by multiple sclerosis. The US trial will involve 250 people and take up to three years to complete. The Wiltshire-based company, G W Pharmaceuticals, also announced that it had obtained a US\$15-million cash infusion from Polygon, an international investment trust.

MARKET WATCH



SOURCE: LUX RESEARCH

Nanotech stocks ended a topsy-turvy 2005 in customary style, with a sharp rise in November followed by a dip last month. The turbulence emphasizes the point that, although investor interest in the fledgling sector seems to be growing, returns are far from assured.

The Lux Nanotech Index tracks about 30 companies: most of them specialize in nanotechnology equipment or applications, but a few are large manufacturing companies that make use of the new technology.

The index fell slightly over the course of 2005 — and ended the year significantly below its peak value of almost 2,000, attained in April 2004.

Nonetheless, Peter Hebert, founder of New York-based consultancy Lux Research, which compiles the index, claims that the upturn in November bodes well for the new year.

Nanotechnology stocks "are starting to

outperform the market and we expect that to continue", he says.

Strong performers towards the end of the year included Westaim, whose stock rose when it said it would stage an initial public offering of shares in its daughter company Nucryst. The Massachusetts-based subsidiary makes wound dressings, based on silver nanoparticles, that fight infection and inflammation.

The offer took place on 22 December and raised US\$45 million. But it provided no cash-in for initial investors, whose shares on the Nasdaq have since remained stubbornly stuck at their opening price of \$10.

Accelrys, a San Diego company that sells software to help others apply nanotechnology, also surged in value after announcing a nanobiology initiative that will be chaired by top biologist Leroy Hood.

Neuroscience gears up for duel on the issue of brain versus deity

SIR — The argument over evolution versus intelligent design, discussed in your News story “Day of judgement for intelligent design” (*Nature* **438**, 267; 2005), is a relatively small-stakes theological issue compared with the potential eruption in neuroscience over the material nature of the mind.

Siding with evolution does not really pose a serious problem for many deeply religious people, because one can easily accept evolution without doubting the existence of a non-material being. On the other hand, the truly radical and still maturing view in the neuroscience community that the mind is entirely the product of the brain presents the ultimate challenge to nearly all religions.

The slow ramping up of this debate, from Descartes’ dualism in the seventeenth century to the neurophilosopher materialists’ claims of victory today, is about to spill over from an esoteric mind–brain debate to the divisive question of whether a product of the mind, such as God, can have any traditionally valid existence whatsoever.

The debate becomes whether a deity, on one hand, stems from human imagination or biological drive or, on the other hand, has an authentic existence that the brain has evolved to perceive.

The reappearance of dualism brings back dusty old memories of long-ago battles that may now need to be refought. As we saw from the media ruckus raised by the Dalai Lama’s address to November’s Society for Neuroscience meeting in Washington DC (even if this did turn down to a rather low simmer on site), the potential for impassioned disagreement exists.

The matter now stands at an intellectual impasse, waiting for an issue around which polarized views will crystallize. We can expect some heady days.

Kenneth S. Kosik

Neuroscience Research Institute,
Department of Molecular and Cellular and
Developmental Biology, Bio II,
University of California, Santa Barbara,
California 93106-5060, USA

Testing is necessary on animals as well as *in vitro*

SIR — Your News Feature “More than a cosmetic change” (*Nature* **438**, 144–146; 2005) includes an emotive photograph showing the heads of six white rabbits, immobilized to have substances dropped into their eyes, with the caption “Tests that put chemicals into the eyes of rabbits have changed little since the 1940s”.

This is not true, at least as far as Britain is concerned. The Home Office, which is responsible for regulating experiments on living animals in Britain, issued guidelines in 1987 for eye irritation/corrosion tests (the Draize test), designed to reduce the pain and injury the test may cause. For example, a substance expected from its chemical nature to be seriously painful must not be tested in this way; the test is permissible only if the substance has already been shown not to cause pain when applied to skin, and *in vitro* pre-screening tests are recommended, such as a test on an isolated and perfused eye. Permission to carry out the test on several animals is given only if the test has been performed on a single animal and a period of 24 hours has been allowed for injury to become evident.

The interesting News Feature in which this photograph appears is unduly dismissive of experiments on living animals. What are the alternatives? The possibilities are either to stop the development of new drugs for human and veterinary use, or to put new drugs on the market without testing them on living animals, or to test new drugs on humans without previous testing on other animals. Few people would be prepared to accept any of these.

To speak of *in vitro* tests as ‘alternatives’ to testing on living animals is misleading: both are necessary. It is impossible to imitate *in vitro* the unimaginable complexity of a human being or indeed of any mammal. *In vitro* tests on bacterial cultures and tissue cultures are necessary in the early stages of testing the very large numbers of substances that are synthesized in order to produce a single drug for use on humans. These tests eliminate all but a very few of those substances, and only those few are candidates for testing on living animals.

Andrew Huxley

Trinity College, Cambridge CB2 1TQ, UK

Animal-rights extremists lose public support

SIR — I am disappointed at the negative tone of your recent News stories “UK animal labs still under siege” (*Nature* **438**, 716; 2005) and “Animal-rights militancy exported to US and Europe” (*Nature* **438**, 717; 2005). The situation in the United Kingdom is indeed serious, but not nearly as grim as you suggest.

UK Home Office figures for 2003 and 2004 show that the number of animal experiments in this country is rising slightly, and since then at least two additional large-scale transgenic-animal laboratories have opened. We are all aware of the setbacks, but the overall picture is hardly that of animal research being driven out. It is true that animal-rights groups have won a few

victories in Britain during the last few years, but they have done so at the cost of alienating the vast majority of the general public. A recent ICM opinion poll (see www.icmresearch.co.uk/reviews/latest-polls.asp) indicates that a clear majority of UK adults support the use of animals for medical research. During the past decade — which has seen animal-rights extremist campaigns of unprecedented scale, ferocity and sophistication — public and media support for the use of animals in research has in fact increased significantly.

The University of Oxford, with the support of the UK government and the overwhelming support of its own student body (www.cherwell.org/show_article.php?id=3868), is willing to face down the extremists.

These are strong indications that the animal-rights extremist campaign has reached its high-water mark and that the tide is now turning against it. The victory against extremism is there for the taking, but the scientific community must learn to reach up and grab it. That means refusing to be intimidated, standing up for our science and, perhaps most important, staying positive.

P. Browne

Address provided

Why should child care be seen as a women’s issue?

SIR — It has been a few decades since families started raising their daughters to have high personal and professional ambitions. But still only a small minority of women remain in science after their postdoctoral phase — mainly those who have postponed or forgone motherhood, or are among the lucky few with access to high-quality affordable child care.

The fact that the issue of childcare availability is discussed in relation to women’s careers (*Nature* **437**, 296 and 446–447; 2005), instead of young scientists’ careers in general, speaks for itself of the bias regarding the role of women in the family and in the workplace.

If women are to share positions that were traditionally occupied almost exclusively by men, what we need is not just affordable child care but a new social ‘family contract’, coherent with expectations about women’s self-fulfilment and the maintenance of the family as an important institution.

Maria José Hötzel

Laboratório de Etologia Aplicada, Departamento de Zootecnia e Desenvolvimento Rural, Universidade Federal de Santa Catarina, Florianópolis, SC 88.034-001, Brazil

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words and should be signed by no more than three authors; preferably by one.

BOOKS & ARTS

In the grey zone

If behaviour arises from interactions between genes and the environment, in what sense is it hardwired?

Hardwired Behavior: What Neuroscience Reveals about Morality

by Laurence R. Tancredi

Cambridge University Press: 2005. 250 pp.
\$28.99**Erik Parens**

We seem beset by an instinct to make binary oppositions: either our choices are free or they are hardwired; we are shaped by either nurture or nature; mental processes are important or physical processes are; morality is a social construct or it is biological in origin. Yet one of the most remarkable things about animals like us is that we can see those binary oppositions for what they are. We can understand why these 'either or's are wholly inadequate for understanding the phenomena at hand.

Laurence Tancredi's book *Hardwired Behavior* powerfully presents science that shows the gross inadequacy of the binary terms we often use to talk about the genesis and character of complex human behaviours. He writes: "Our brain structures are not immutable; they are susceptible to change for the better and change for the worse." Indeed, much of the research he discusses rests on this neuroplasticity. He reports on research showing that talk therapy can produce neuronal changes. His chapter on gender differences suggests that changing social conceptions of the roles of women "will inevitably affect the biology of their brains over time". He reports on research showing that rats deprived of nurture at birth fail to express a gene that is correlated with their ability to handle stress. And he refers several times to a fascinating study by Avshalom Caspi and colleagues (*Science* **301**, 386–389; 2002), which found that the likelihood of children becoming antisocial as adults is a function of both their genomes and their experiences. As Tancredi observes, this finding "emphasizes the interactive nature of genes and environment, nature and nurture".

Why then did Tancredi call his book *Hardwired Behavior*? Because he wants to underscore that those interactions between genes and the environment result in brains with strong dispositions. He also wants to convey the excitement surrounding scientific and technological advances that enable researchers to glimpse the neural correlates of those dispositions. Moreover, as a lawyer and a psychiatrist, he wants to show that the more we understand

the neurobiology of those dispositions, the less we should believe that individuals are as free to choose as judges and moralists may assume.

So, for example, he describes the horrific childhood environment of the murderer Ricky Green and, invoking the Caspi study, speculates about Green's genotype. His primary point is that Green's experience altered the biology of his brain. He wants us to appreciate the "deeply physical nature" of Green's dispositions as an adult. This point is in a sense more profound than it may seem. If one believed that human choices originate in a metaphysical soul, in an entity beyond nature, then it would come as a great surprise to learn that human choices emerge out of staggeringly complex interactions within and between biological and social systems. But if one long ago gave up the idea of a metaphysical soul, the point about the deeply physical nature of our behaviours seems less profound and is potentially misleading.

Consider Tancredi's account of his patient Art, whose "biologically compelled" sexual escapades destroyed his relationship with his fiancée. After the break-up and several months of psychotherapy, Art entered a new relationship. He admitted that he still had the urge to get involved with other women, but this time around he could admit that he had these feelings. "He knew it would be a long and at times rocky road to having a truly successful relationship, but he was now committed to make this happen."

Art's sexual escapades and his commitment to stop them are both deeply physical in the broad sense that he couldn't experience them if his genes and neurons weren't working exactly as they were. But leaving aside rare cases where a single genetic mutation or brain lesion causes an aberrant behaviour, much less is known about the causal pathways that lead to complex human behaviours than this "deeply physical" language seems to suggest. Moreover, when more is known about these causal pathways, the essential elements will be not just genes and neurons, but also words, human relationships and social customs.

To the extent that this "deeply physical" language seduces us into thinking that our behaviour is hardwired — in the sense of being determined by genes and neurons alone — it is deeply unhelpful. To the extent that it reminds us that our choices don't arise in the way the metaphysical model suggests, it is helpful. And to the extent that it is an expression of the desire to move from condemning bad behaviours to understanding them, it is deeply generous.

Thinking entails noticing when we have allowed the instinct for binaries to keep us from the complexity of the phenomena at hand. In much of *Hardwired Behavior*, Tancredi helps us get over that instinct, and for that we should be grateful. ■

Erik Parens is at The Hastings Center, 21 Malcolm Gordon Drive, Garrison, New York 10524-5555, USA.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Slaves to biology: do our genes and the environment remove our freedom to control our own behaviour?

The monster that is medicine

Dr Golem: How to Think about Medicine

by Harry Collins & Trevor Pinch

University of Chicago Press: 2005. 280pp.
\$25

W. F. Bynum

Having told us how to think about science (*The Golem*, Cambridge University Press, 1993) and technology (*The Golem at Large*, Cambridge University Press, 1998), Harry Collins and Trevor Pinch turn their attention to an even softer target: medicine. Readers of the previous two books will be familiar with the structure: a series of case studies, each based on a couple of principal sources, with linking commentary. They will also know that the series takes its title from a creature of Jewish mythology. Golems were powerful man-made beings that would follow orders, but they were a bit thick and did not know their own strength, and so were potentially dangerous. Collins and Pinch have never sought to teach their readers science, technology or medicine — merely how to think about these complex subjects, each of which, they argue, has golem-like characteristics.

For medicine, they invite us to consider eight topics, each with wider implications for the current state of medicine, both as science and as art (or, as they prefer to call it, succour). The translation of medical science into clinical art is rarely easy. For one thing, most medical practitioners, although trained in science, are not scientists, nor do their daily duties encourage them to be. In addition, the placebo effect — much studied but not easily explained by science — colours much of medical practice and both doctors' and patients' unconscious

evaluation of their encounters. Collins and Pinch argue that the placebo effect is the hole in the heart of medicine, making the assessment of therapeutic interactions and the introduction of new drugs problematic.

Later chapters describe the difficulty of uncovering bogus doctors (those practising without qualifications); the diagnostic variability among 'real' doctors when confronted with such things as enlarged and inflamed tonsils; the problem of contested diseases such as chronic fatigue syndrome, Gulf War syndrome and fibromyalgia; the dubious effectiveness (but sacrosanct position) of cardiopulmonary resuscitation; and the current debate about the relationship between autism and vaccinations, especially the triple measles, mumps and rubella (MMR) vaccine. A chapter on HIV and the patient as expert is reprinted from the earlier volume on technology. What is at stake with each of these topics is the fundamental uncertainty of medicine as science and its clumsiness as succour.

The authors write as patients as well as sociologists, and the chapter on MMR, on which subject Collins and Pinch have diametrically opposed attitudes, is particularly effective. It has to be said that the authors want medicine to be more scientific, not less so. Their volume is hardly a plea for alternative medicine or even for the uncritical democratization of conventional medicine. Rather, it is an analysis of the problems of contemporary medical knowledge, and is stronger on diagnosis than on prescription. The authors offer stringent critiques of modern medicine's inadequacies, but are reluctant to suggest what might be done to change it for the better.

This detachment was also present in the earlier volumes and stems, I think, from their attitude towards the nature of 'expertise'. Their discussion of HIV makes a strong case for sufferers having their own kind of expertise, and this may not be simply about the subjective nature of their illness. Many AIDS activists became expert in the nuances of retroviruses, how antiviral drugs work and the design of clinical trials. But as Collins and Pinch point out, meaningful dialogue about scientific issues requires that both parties know what they are talking about. The Internet may not be the best place to acquire expertise, for its unregulated nature most starkly exposes the difficulties of democratic knowledge.

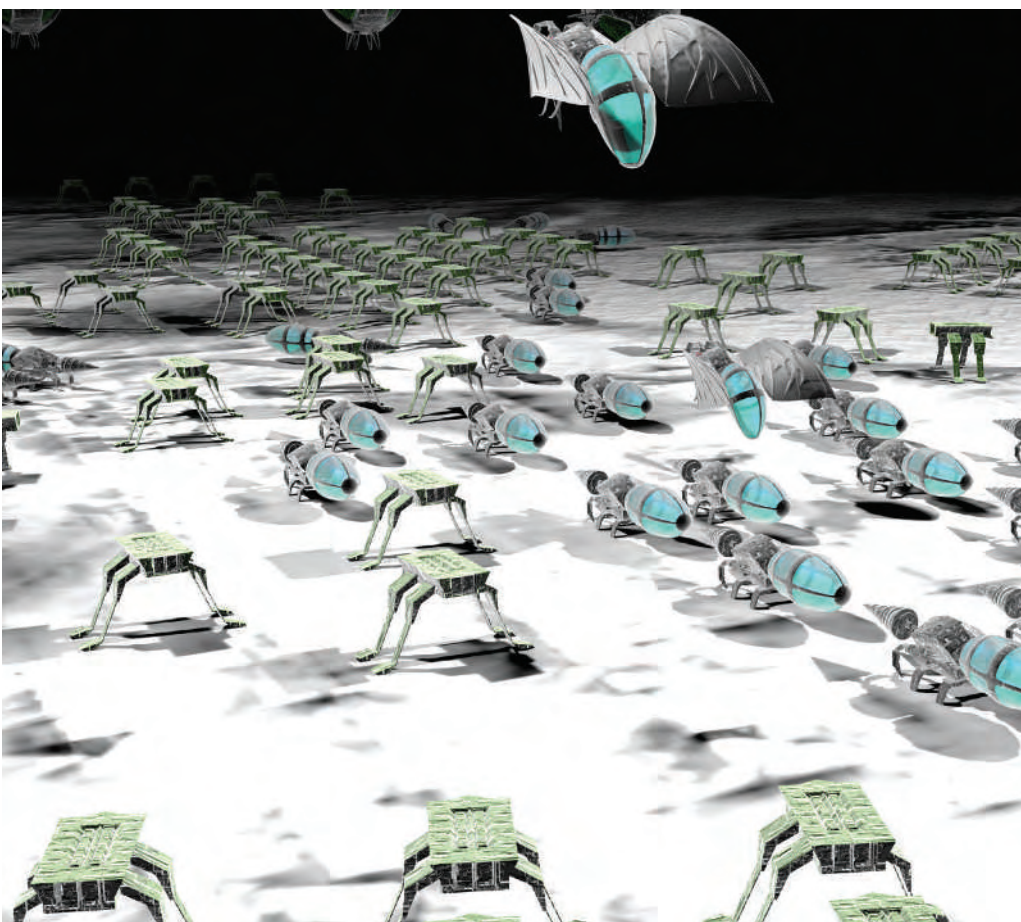
The authors have selected their topics reasonably well, but there are some gaps and curious decisions. The chapter entitled 'Alternative Medicine' is not about alternative medicine at all. It looks at Linus Pauling and Ewan Cameron's advocacy of massive doses of vitamin C as a treatment for cancer. Pauling and Cameron worked entirely within the framework of scientific medicine, offered a testable hypothesis about why vitamin C might have the effect they postulated, and wanted clinical trials to be done. They disagreed about the actual design and implementation of the trials, and of course about their outcome, but this episode is about scientific, not 'alternative' medicine.

Collins and Pinch's familiar, cosy style may grate on some, but their purpose is high-minded. They have in fact discovered what the Hippocratics knew more than two millennia ago: "Life is short and art long; the occasion fleeting; experience fallacious, and judgement difficult."

W. F. Bynum is at the Wellcome Trust Centre for the History of Medicine, University College London, 210 Euston Road, London NW1 2BE, UK.

Operating differently:
medicine deviates
from science in many
ways, not least by
embracing the
placebo effect.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS



Never mind the hype: weighing up nanotechnology requires an understanding of why people are afraid.

A little judgement

Nano-Hype: The Truth Behind the Nanotechnology Buzz

by David M. Berube

Prometheus: 2005. 300 pp. \$28

Harry Collins

Major technologies are born into a febrile world these days, with everyone wanting a piece of the action. Where projects are so large that the costs fall on wide sectors of society, the decision-makers are rarely technical experts. Giving an idea a sexy label attracts the attention of the powerful, irrespective of its technical merit or real promise. Consider the billions that were poured into the 'third generation' of intelligent computers, the space race and the search for a 'magic bullet' for cancer.

Nowadays, the public, led by pressure groups and represented by politicians, believe that they too have a right to a say regarding the implications of a new technology, especially with respect to its dangers and its impact on the world. It is a claim that has been hard to deny ever since it became clear that the "power too cheap to meter", promised after the Second World War, was a bluff. Genetic engineering and the combined measles, mumps and rubella (MMR) vaccine have paid the price.

But what about nanotechnology, another player in this game? Should we be optimistic about its potential, or is pursuing it an obvious mistake? This is the question posed by David Berube in his book *Nano-Hype*.

In practice, it is often difficult to see how to make such a judgement about the value of a technology independent of the overblown optimism and overly dramatized foreboding. Sometimes, though, it is relatively easy. The fuss over the MMR vaccine, for example, should never even have started because there was no scientific controversy — there was no evidence at all that MMR was dangerous, even though the newspapers failed to grasp the point. On the other hand, in the 1950s the UK media sang the praises of Britain's first nuclear power station, Calder Hall, but it was scandalous that it took so long for it to be understood that the cost of decommissioning nuclear plants should be included in the price of the power. These cases are rare because they are so clear cut; there are usually pretty good arguments on both sides.

One way to make progress in difficult cases would be to analyse the mechanisms that amplify the arguments on either side of the debate. Perhaps this would make it possible to

discount some of the fuss, even if it wouldn't necessarily yield a neat resolution. How do the advertising campaigns and the associated moral panics work? Who pumps them up and why? What are the roles of the money-hungry researchers and their institutions, and the media? To find out how positive and negative hype works, you have to look at detailed case studies of specific incidents and comparative analysis of the many well known episodes.

Unfortunately, Berube's book does not take this approach. The author makes the mistake of thinking that a sufficiently exhaustive look at every word that has ever been written about nanotechnology will reveal something. But science is an oral culture. Although science's spokespersons rattle on endlessly about peer review, the vast majority of published papers, peer reviewed or not, are largely ignored by scientists in the field. The problem that would face an alien from another planet who wanted to make a digest of terrestrial science from the literature alone would be about as bad as that facing a lay person who tries to understand it by reading everything on the Internet.

To know what counts as reliable science you need to have the written word sifted by someone who knows how to sift it — someone who is embedded in the oral culture that selects and condenses those scientific contributions that are of lasting importance. There is usually more than one way to do the selection, and this is why experts disagree, but without any sifting there is nothing of substance to argue about. To sort out the material that contributes to a scientific controversy, then, one must either be a scientist or spend a long time talking to the specialists and pulling out the major themes around which disagreements and misunderstandings turn.

Berube has instead produced a long, closely printed and almost unreadable text that quotes from a huge range of sources, primary and secondary, with no winnowing with respect to value. It is organized on the principle of a list, rather than on the explanatory hypotheses. Interspersed in the text are some disconnected musings. To give one example, here is Berube's 'explanation' of why Richard Feynman became interested in nanotechnology: "What led to Feynman's lectures on microtechnology? According to [Freeman] Dyson, Feynman struggled 'to understand the workings of nature by rebuilding physics from the bottom up.' Hence, it might not be much of a stretch for Feynman to transpose his bottom-up view of the discipline of physics to fabrication strategies of the very small." There is not much one can do with this kind of thing.

But because the referencing is so dense, at least the book provides a useful pointer to the available secondary source material on nanotechnology. ■

Harry Collins is at the Centre for the Study of Knowledge, Expertise and Science, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff CF10 3WT, UK.

A bigger picture of apes

The recent *King Kong* film highlights how our perceptions of gorillas have changed.



Janet Browne

In one way or another, apes have always been in show business. They have long been a source of spectacle, wonder, horror and, as often as not, moral parables about ourselves. With the new version of the film *King Kong*, the producers have tapped into a rich, almost primal, experience. An adventure story full of classic metaphors and stunning visual effects, it pays generous homage to the original 1933 film, while intensifying and filling out the drama.

Kong is much more humanized than in the original, a change that evidently reflects the requirements of modern audiences, as well as a greater public knowledge of gorillas, thanks to wildlife programmes and widespread publicity for conservation issues. What's more, Kong's blonde (Naomi Watts in the new film) now falls in love with him, a material departure from the original that responds to the idea that gorillas are our closest evolutionary relatives.

Yet he is more gorilla than before, too. The first glimpse shows Kong as a mature silverback, and his movements and behaviour are incredibly naturalistic. A scene in Kong's lair shows the skeletons of other giant apes, a suggestion that he is the last of his kind — fierce only because he is lonely. The final phrase "It was beauty killed the beast" movingly acknowledges the opposing forces of nature and mankind, and the fragile bonds between them.

Anthropomorphisms like these are

intriguing to cultural historians. Gorillas were unknown to Western naturalists until the 1850s when mysterious bones were sent back from Africa and received a scientific name. Humans thereafter regularly projected their own assumptions on to the species, and gorillas soon came to represent the dark side of human existence.

Even though ape ancestors were not mentioned by Darwin in *On the Origin of Species*, they erupted into the evolutionary debate with the first displays of stuffed specimens in Europe promoted by Paul du Chaillu in 1861. Du Chaillu's stories of the gorilla's ferocity were at once terrifying, amusing and revolting to most Victorians and are often cited as an early source for

King Kong. One of those vintage specimens is preserved in Melbourne, Australia.

With the rise of the great museums in the early twentieth century, interest in capturing and filming live animals led pioneering curators such as Carl Akeley of the American Museum of Natural History to mount lavish collecting expeditions. Akeley created a remarkable display that served as a starting place for Willis O'Brien, the model maker for the original *King Kong*. Indeed, Akeley also humanized gorillas, regarding himself as an old male silverback, and established the first nature reserve for them in Rwanda, where he was buried.

At the same time ethologists tended to see warlike, competitive human society replicated in apes. The notion of dominant alpha males lasted for more than 40 years until George Schaller and Dian Fossey observed gorilla behaviour in the wild and showed them living cooperatively in small family groups. Television, film and modern ecotourism have recently presented gorillas to the public as shy and friendly animals, a mirror image of how we would like to see ourselves. The original *King Kong* of the 1930s stood at the crossroads. This new film invites us with wit, insight and great moments of heroic imagery to reflect on what it is to be human. ■

Janet Browne is at the Wellcome Trust Centre for the History of Medicine, University College London, 210 Euston Road, London NW1 2BE, UK. Browne, Jordan Goodman and Robert Peck are currently writing a cultural history of the gorilla.



Early views of gorillas were very different to the more realistic ape in *King Kong*.

NEWS & VIEWS

EXTINCTIONS

A message from the frogs

Andrew R. Blaustein and Andy Dobson

The harlequin frogs of tropical America are at the sharp end of climate change. About two-thirds of their species have died out, and altered patterns of infection because of changes in temperature seem to be the cause.

One of the worries about global climate change is that it will raise the transmission rates of infectious diseases¹. On page 161 of this issue, Pounds and colleagues² provide compelling evidence that anthropogenic climate change has already altered transmission of a pathogen that affects amphibians, leading to widespread population declines and extinctions.

According to the Global Amphibian Assessment (GAA)³, around a third of amphibian species (1,856) are classified globally as 'threatened'. The tenuous hold these animals have on life is especially evident in tropical America, where, for example, 67% of the 110 species of harlequin frog (*Atelopus*; Fig. 1) endemic to the region have died out in the past 20 years³. A pathogenic chytrid fungus, *Batrachochytrium dendrobatidis*, is implicated as the primary cause of *Atelopus* population crashes and species extinctions^{4,5}. Now, Pounds *et al.* offer a mechanistic explanation of how climate change encourages outbreaks of *B. dendrobatidis* in the mountainous regions of Central and South America: night-time temperatures in these areas are shifting closer to the thermal optimum of *B. dendrobatidis*, and increased daytime cloudiness prevents frogs from finding 'thermal refuges' from the pathogen.

The authors defined an 'extinction' as the time when a frog species was last observed by professional teams of herpetologists working in these regions. Most extinctions (78–83%) occurred in years that were unusually warm across the tropics. The likelihood that this correlation arose by chance is less than one in a thousand.

Moreover, the observed patterns of extinction vary with altitude — as do the effects of climate change. Montane *Atelopus* species that live between 1,000 and 2,400 metres show higher rates of extinction than do those that live only in the lowlands (where extinctions are rare) or just in the highest elevations. Pounds *et al.* propose that this is because the extreme sites afford thermal refuges, with temperatures being either too high or too low for

optimal growth of the pathogen. Mid-elevation *Atelopus* communities are not only the hardest hit by extinction, but they also harbour the most species, so biodiversity in these areas is in double jeopardy. These results corroborate the GAA findings³ for a broad array of amphibians that the percentage of extinct or threatened species is largest at middle elevations. This is contrary to the expectation that higher-elevation species would be more prone

change had been stymied by the so-called 'climate–chytrid paradox', because the climatic conditions favouring chytrid growth seemed to be the very opposite of those created by current climate trends.

Pounds and colleagues' work² is a breakthrough as it resolves the paradox and offers a theory to explain the widespread 'enigmatic' declines of *Atelopus* and other amphibians³. The authors combine two disparate approaches

into one unifying theory, simultaneously explaining how shifting temperatures are the ultimate trigger for the expansion of a pathogenic fungus, and that this infection is the direct cause of *Atelopus* extinctions.

There may be a tragic irony here. The oldest-known hosts of *Batrachochytrium* are African-clawed frogs (*Xenopus*)⁷, first recorded in South Africa in 1938. Global trade in these frogs burgeoned in the 1950s following the development of pregnancy tests that used *Xenopus* tissue^{7,8}. Museum records suggest that the pathogen achieved a worldwide distribution in the 1960s. So it seems that the expansion in one frog species through trade may have led to the extinction of other amphibian species — a totally unexpected, indirect consequence of human ingenuity.

Frogs and *Batrachochytrium* fungi are not the only example of synergistic interactions between pathogens and climate change that are affecting biodiversity. The climate change in the Arctic and sub-Arctic has modified the life cycle of the nematode parasites of musk oxen⁹. These worms can now complete their life cycle in one year, instead of two, and their rising numbers are having a significant impact on the survival and fecundity of musk oxen.

Similarly, warmer climate conditions in montane regions in the western United States allow the mountain pine beetle (*Dendroctonus ponderosae*) to complete its life cycle in one year, rather than two. These beetles transmit pine blister rust (*Cronartium ribicola*), and as they become more abundant, the fungus they



Figure 1 | Amphibian alarm call. The Panamanian golden frog is one of roughly 110 species of harlequin frog (*Atelopus*), many of which are dying out. Although this species still survives, its numbers have fallen significantly.

to extinction because they generally have smaller environmental ranges over which they can survive.

Although the little-known *Batrachochytrium* fungus was proposed to be potentially the sole reason for declines in amphibian populations in the tropics^{4,5}, no one had come up with an explanation for the sudden emergence of this pathogen. Moreover, although chytrid disease was a common condition in many areas experiencing declines, it was not clear whether *Batrachochytrium* was directly responsible or whether the infection was a secondary effect associated with dead or dying animals⁶. Previous attempts to explain the prevalence of the disease in terms of climate

F. BREVI

carry is beginning to have a serious effect on the pine trees in the highest-elevation forests along the Rocky Mountains^{10,11}. Like *Batrachochytrium*, pine blister rust has the potential to eradicate several host species, so it could lower tree lines throughout the Rockies and cause increased run-off and flooding.

The powerful synergy between pathogen transmission and climate change should give us cause for concern about human health in a warmer world^{12,13}. The ubiquity, complexity and cascading effects of host–pathogen interactions make their dynamics extremely difficult to predict. As global change is occurring at an unprecedented pace, we should expect many other host taxa, from ants to zebras, to be confronted with challenges similar to those faced by *Atelopus*. We should also expect the unexpected: terms such as ‘enigmatic decline’ and ‘pathogen–climate paradox’ will probably dominate explanations of extinctions until we develop a better understanding of the relationships between global change, pathogens and their hosts. Few of the current models and assessments of biodiversity that are used to forecast extinctions or identify taxa at risk include information on how climate affects disease dynamics. Until they do, they will enjoy limited success and will probably give overly optimistic prognoses of how biodiversity will

be affected by climate change. The frogs are sending an alarm call to all concerned about the future of biodiversity and the need to protect the greatest of all open-access resources — the atmosphere. ■

Andrew R. Blaustein is in the Department of Zoology, 3029 Cordley Hall, Oregon State University, Corvallis, Oregon 97331-2914, USA. e-mail: blaustea@science.oregonstate.edu

Andy Dobson is in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544-1003, USA. e-mail: andy@eno.princeton.edu

1. Dobson, A. P. & Carper, E. R. *Lancet* **342**, 1096–1099 (1993).
2. Pounds, J. A. et al. *Nature* **439**, 161–167 (2006).
3. www.globalamphibians.org (2004).
4. Daszak, P. et al. *Emerg. Infect. Dis.* **5**, 735–748 (1999).
5. Berger, L. et al. *Proc. Natl Acad. Sci. USA* **95**, 9031–9036 (1998).
6. McCallum, H. *Conserv. Biol.* **19**, 1421–1430 (2005).
7. Ouellet, M. et al. *Conserv. Biol.* **19**, 1431–1440 (2005).
8. Rachowicz, L. J. et al. *Conserv. Biol.* **19**, 1441–1448 (2005).
9. Kutz, S. J. et al. *Proc. R. Soc. Lond. B* doi:10.1098/rspb.2005.3285 (2005).
10. Kendall, C. K. & Roberts, E. K. in *Whitebark Pine Communities: Ecology and Restoration* (eds Tomback, D. F., Arno, S. E. & Keane, R. E.) 221–242 (Island Press, Washington DC, 2001).
11. Logan, J. A. & Powell, J. A. *Am. Entomol.* **47**, 160–172 (2001).
12. Dobson, A. P., Kutz, S., Pascual, M. & Winfree, R. in *Climate Change and Biodiversity: Synergistic Impacts* (eds Lovejoy, T. & Hannah, L.) (Yale Univ. Press, 2003).
13. Patz, J. A. et al. *Nature* **438**, 310–317 (2005).

SPACE PHYSICS

Breaking through the lines

Götz Paschmann

Magnetic field lines are known to reorganize themselves in plasmas, converting magnetic to particle energy. Evidence harvested from the solar wind implies that the scale of the effect is larger than was thought.

The reconnection of magnetic fields that occurs in the ionized gases known as plasmas is a fascinating and enigmatic phenomenon. It transforms magnetic field configurations, converting energy stored in those fields into kinetic energy of the electrically charged particles that make up the plasma. Direct and indirect proof for the existence of the effect comes from many different quarters, ranging from explosive energy releases in the Sun's atmosphere to catastrophic disruptions in nuclear fusion reactors. On page 175 of this issue, Phan *et al.*¹ present the latest observations of magnetic reconnection in the solar wind — a stream of plasma that is continuously emitted by the Sun — and in doing so clarify the spatial and temporal scales that govern the process.

Plasmas, which in space consist mainly of protons and electrons, are commonly permeated by magnetic fields. Plasma and field tend to behave as if frozen together: the plasma's particles lend magnetic field lines physical

form by gyrating around them; equally, when the particles move, the magnetic field lines move with them. This means that, rather like an individual strand in a bowl of spaghetti that is being stirred, the same field line — although constantly changing position and shape — always connects the same particles of the plasma (Fig. 1a).

But what happens if the plasma's motion brings together two magnetic field lines that point in opposite directions? The frozen-in picture assures us that all particles will remain on their respective field lines, regardless of how hard these are pushed together. But this picture is only an approximation, and in some circumstances — poorly understood at present — field lines slip relative to the plasma, and break and cross-link at an ‘X-point’ (Fig. 1b). The field lines, now sharply bent, act as a slingshot, imparting their stored energy to the particles and ejecting them at high speeds. This is the phenomenon known as magnetic reconnection^{2–4}.

Field lines in neighbouring planes may also reconnect, resulting in an ‘X-line’ linking many X-points (Fig. 1c). This is essentially what happens when the solar wind encounters Earth's magnetic field. The solar wind transports the Sun's magnetic field into interplanetary space, so it could not penetrate Earth's field if the frozen-in theorem held. But direct and indirect observations^{5–7} show that the frozen-in condition breaks down on the magnetopause, the surface that separates the solar wind and Earth's magnetic field. This allows terrestrial magnetic field lines to become connected directly with the Sun, and so the solar-wind plasma flows along reconnected field lines into Earth's magnetosphere.

But can reconnection happen within the solar wind itself? Because of the complex spatial and temporal variations in conditions near the Sun's surface, there are abrupt changes in the density and velocity of the solar wind, and associated rotations in the direction of its magnetic field. Without the breakdown of the frozen-in theorem, the plasmas on the two sides of such a transition would never mix. If such a breakdown occurs, so, by necessity, does magnetic reconnection: a spacecraft positioned in the solar wind would see the passage of a transition not just through sudden changes in the properties of the plasma and magnetic field, but through high-speed plasma flows characteristic of reconnection. NASA's Advanced Composition Explorer, ACE, has recently observed exactly this^{8,9}.

The ACE observations used a single spacecraft, so ACE could not measure the length of the X-line causing the plasma flows. It was also not clear whether reconnection was active for longer than the few minutes it took the reconnection layer to sweep over the spacecraft at solar-wind speed. Phan *et al.*¹ address these open questions by taking advantage of a fortuitous configuration of three spacecraft — NASA's ACE and Wind, and one of the European Space Agency's four Cluster spacecraft — that gave the researchers a large baseline for their measurement (see Fig. 1 on page 175). On 2 February 2002, one after the other, all three spacecraft recorded the passage of a reconnection layer with essentially identical characteristics, in particular the same net plasma and magnetic field changes and the same plasma flows.

The observed plasma flows agreed quantitatively with theoretical predictions based on the change in magnetic field across the layer and the local plasma density. Once the authors had inferred the direction of the X-line from a simple geometrical argument, they could calculate that the X-line must have been at least 2.5 million kilometres long — almost 200 times the diameter of Earth. And from the spacing of the passage times over the three spacecraft, it was evident that reconnection was not explosive, but instead operated steadily for at least two-and-a-half hours.

Phan and colleagues' observations¹ of the

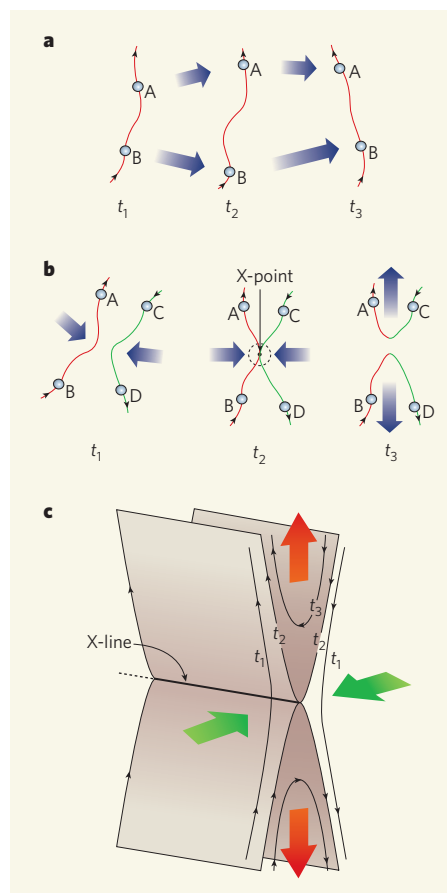


Figure 1 | The ins and outs of magnetic reconnection. **a**, Magnetic field lines are generally frozen into the plasma flow (blue arrows), so two charged particles, A and B, connected by a field line at time t_1 , remain connected by the same field line at all later times. **b**, Two oppositely directed field lines, identified by particles A, B and C, D, respectively, are moving towards each other at time t_1 . When they touch at time t_2 , they break and cross-link (reconnect) at the so-called X-point, leaving A, C and B, D, connected at time t_3 . The highly bent field lines act like a slingshot, and plasma flows out from the region at high speeds. **c**, A perspective view of magnetic field lines reconnecting along an X-line (plasma inflow, green arrows; high-speed plasma outflow, red arrows). The times t_1 to t_3 refer to the same phases of the process as in **b**. Phan *et al.*¹ investigate the length of the X-line in the solar wind.

spatial and temporal characteristics of magnetic reconnection will fuel the sometimes heated debate over what the phenomenon is like and what it can do. With the launch of NASA's STEREO mission, expected later this year, larger baselines will become available, and there is hope that the study of solar-wind reconnection can be extended to much larger scales. At the smallest scales, NASA's Magnetospheric Multi-Scale (MMS) mission, to be launched in 2013, will investigate the kinetic plasma processes near the X-line that allow the frozen field condition to be broken and reconnection to occur. The prospects for finally understanding the nature of reconnection, its ability to couple small- to large-scale phenomena, and the

crucial role it plays in various cosmic settings, are excellent.

Götz Paschmann is at the Max-Planck-Institut für extraterrestrische Physik, 85748 Garching, Germany.

e-mail: goetz.paschmann@mpe.mpg.de

1. Phan, T. D. *et al.* *Nature* **439**, 175–178 (2006).
2. Dungey, J. W. *Phys. Rev. Lett.* **6**, 47–48 (1961).

3. Petschek, H. E. in *AAS-NASA Symp. Physics of Solar Flares* (ed. Hess, W. N.) 425–439 (NASA, Washington DC, 1964).
4. Sonnerup, B. U. Ö. *J. Plasma Phys.* **4**, 161–174 (1970).
5. Paschmann, G. *et al.* *Nature* **282**, 243–246 (1979).
6. Sibeck, D. G. *et al.* *Space Sci. Rev.* **88**, 207–283 (1999).
7. Phan, T. D. *et al.* *Space Sci. Rev.* **118**, 367–424 (2005).
8. Gosling, J. T. *et al.* *J. Geophys. Res.* **110**, A01107 (2005).
9. Gosling, J. T. *et al.* *Geophys. Res. Lett.* **32**, L05105 (2005).

MEDICINE

Politic stem cells

Irving L. Weissman

Research on embryonic stem cells holds huge promise for understanding and treating disease. Many people oppose such research on religious and ethical grounds, but two new methods may bypass some of these objections.

In this issue are two new methods^{1,2} for producing pluripotent stem-cell lines — the great future hope of regenerative medicine*. Both papers report proof-of-principle tests in mice of techniques that might be used for making human pluripotent stem-cell lines. The protocols each aim to satisfy the religious, ethical and/or political objections of groups that are opposed to some of the methods used in embryonic stem-cell research.

Pluripotent stem-cell lines come from the most primitive cells in vertebrate development. They are prized because they can both renew themselves continuously in culture and, once released from this self-renewal cycle, can go on to form most mature cell types in the body (hence 'pluripotent', meaning many potentials). Their ability to make a range of functional cell types makes them crucial to the study of tissue development and degenerative diseases, and they are considered to be promising as a possible treatment for such disorders.

Pluripotent stem-cell lines can be derived from early embryos before they implant in the uterus (Fig. 1a, overleaf). These cells are called embryonic stem cells (or ES cells). The preimplantation embryo (a blastocyst) has an outer shell of cells used for uterus implantation (the trophectoderm) and an inner cell mass of pluripotent cells that will give rise to the developing embryo once it has implanted. To create ES cell lines, cells from the inner mass are removed and cultured, but this process means that the embryo cannot implant in the uterus.

To get around this, Lanza and colleagues¹ (page 216) have adapted a method commonly used in assisted-reproduction clinics for preimplantation genetic diagnosis. This involves removing a cell from the eight-cell stage of development (before the blastocyst

has formed) (Fig. 1b); this 'blastomere' cell is then analysed for genetic defects. Instead, Lanza and colleagues¹ use the blastomere cell to produce ES cell lines — without compromising the embryo from which the blastomere was obtained. The single blastomeres are co-cultured with established ES cell lines, and then separated from them to form fully competent ES cell lines.

The ES cells produced using Lanza and colleagues' technique would have the same genes as the embryo, essentially a mix from the two parents undergoing *in vitro* fertilization treatment. However, the goal for many researchers is to be able to produce pluripotent cells that represent the full genetic diversity of humans, or that are genetically identical to a particular donor (a patient with a genetic disorder, for example). The production of such stem-cell lines would enable the study of the cellular and genetic bases of disease development³. For example, stem-cell lines generated from mice that are immunodeficient because of a defect in a single gene are themselves immunodeficient, and this might hold true for complex multigene disorders such as amyotrophic lateral sclerosis. These lines might also be 'fixed' in culture by replacing the defective gene with healthy copies, and thereby one could validate the role of particular drug targets or the efficacy of certain therapies. In the long term, healthy cells derived from repaired stem cells might aid the regeneration of tissues from the donor patient.

At present, creating pluripotent cells from a specified donor can only be achieved by a process called nuclear transfer (NT) (Fig. 1c). This involves removing the nucleus from a donor body cell (say, a skin cell) and injecting it into an egg that has had its own chromosomes removed. The egg is then encouraged to form an embryo-like, or embryoid, blastocyst, during which process the body-cell nucleus undergoes 'reprogramming', changing from expressing skin genes to expressing more

*The two papers concerned^{1,2} and this article were published online on 16 October 2005. Since then, ref. 4 by W. S. Hwang *et al.* has been brought into question, and its authors have requested retraction of the paper.

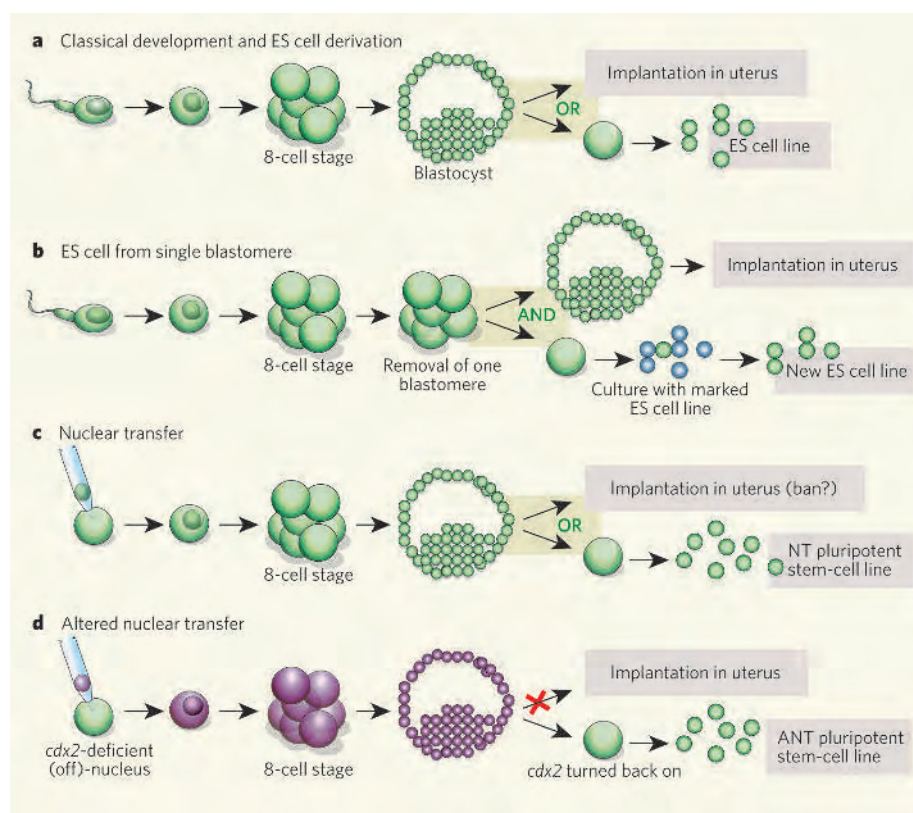


Figure 1 | Producing pluripotent stem-cell lines. **a**, The classical derivation of embryonic stem (ES) cells destroys the embryo from which they are derived. **b**, Lanza and colleagues¹ have used a modified method that does not compromise the embryo, but is not donor-specific. **c**, Donor-specific pluripotent stem cells can be made using nuclear transfer (NT) techniques. **d**, An altered nuclear transfer (ANT) method developed by Meissner and Jaenisch² blocks expression of the *cdx2* gene until the blastocyst stage, making it unable to implant.

pluripotent genes. Embryoid blastocysts have an inner cell mass like normal blastocysts, and these cells can become pluripotent stem cells. Such NT stem cells can, like ES cells, self-renew or differentiate to become most types of mature body cell. Technological advances⁴ have improved NT in humans to the point that a single egg donor can produce enough eggs in one round of donation to ensure a patient-specific NT pluripotent stem-cell line.

Very rarely, animal embryoid blastocysts have reprogrammed enough genes to be able to implant in the uterus and complete all the developmental stages to birth. But in all species tested, more than 99% of embryoid blastocysts fail, many at later stages of pregnancy where the failure can injure or kill the mother bearing it. This has led a panel of the US National Academies to call for a legally enforceable ban on human 'reproductive cloning'⁵. Nonetheless, because human NT stem cells come from embryoid blastocysts, their derivation has raised objections on political, ethical and religious grounds. A possible solution to the controversy, proposed by many who want the medical science to progress, might be to invent a process that produces an entity that cannot implant in the uterus — termed alternative nuclear transfer (ANT) by William Hurlbut, a member of President Bush's Council on Bioethics⁶.

Meissner and Jaenisch² (page 212) have now developed a method to accomplish ANT. Their technique builds on previous work by Strumpf *et al.*⁷, who studied a gene called *cdx2* and its role in establishing the mouse trophoderm and, later, the intestinal tract. Their results suggested that if this gene was suppressed in the nucleus of the donor cell during the NT process, it might allow the generation of NT entities that could not implant.

Meissner and Jaenisch demonstrate that this is indeed the case, using a clever method to control *cdx2* expression at various stages. They introduced into the donor cell a gene encoding an RNA that inhibits *cdx2* expression, and this gene was transmitted with the donor nucleus to the egg and continued to be active during the NT. Once they had derived the ANT pluripotent stem-cell line from the resulting embryoid blastocysts, they clipped out the inhibitor gene to enable the resulting ANT stem cells to produce mature intestinal epithelia given the right cues (Fig. 1d). These ANT pluripotent stem-cell lines can form many other mature cells, just as the classical ES and NT cell lines do.

It is highly speculative whether either blastomere-derived ES cell lines¹ or ANT pluripotent stem-cell lines² can also be derived from human cell sources. Nonetheless, there have already been hearings in the US Congress at



50 YEARS AGO

"Physiological control of population growth" — As Dr. Gregory Pincus, of the Worcester Foundation for Experimental Biology, pointed out, there is no doubt that progesterone can inhibit ovulation in rabbits and apparently, also, in rats. According to his own studies, the indications are that progesterone, when taken by mouth, will also inhibit ovulation in women, as determined by various indirect indices. This view was not, however, shared by Dr. Massomi Ishikawa ... nor by Dr. A. Stone ... [as was clear from all the physiological papers] the practical goal of these urgently needed researches — the discovery of a 'pill' which can be taken by mouth, and the only physiological effect of which would be that of inhibiting the development of the fertilized ovum, or of suppressing ovulation or gametogenesis at will — is so remote from realization that at this stage no one can say how, when or even whether success will ever be achieved.

Sir Solly Zuckerman

From *Nature* 14 January 1956.

100 YEARS AGO

"The training of the body and mind" — In the afternoon Sir Lauder Brunton took the chair, and discussed education in connection with the threefold character of man. At first, he said, moral training was provided, and churches and cathedrals were built long before the people could read or write; then mental culture was considered, and became very general; and, lastly, it was being recognised that the condition of the body had considerable effect upon the morals and the mind, so that a physical training was also considered necessary. He gave some interesting instances to show how character and habits had been entirely altered by accidents to the brain, and said that while Newton was physically weak, Young, who was his superior, even in mental capacity, was a circus rider, and could perform almost any bodily feat.

From *Nature* 11 January 1906.

50 & 100 YEARS AGO

which some representatives called for a moratorium on the production of further stem-cell lines until these methods work in humans, while others declared that medical science has already gone too far, and must be reined in by laws that criminalize all such attempts. Of course, the attempts to delay or to prevent these kinds of experiment derive from the belief that preimplantation embryos, or entities with little or no potential to form a functioning organism, are human, and have the same rights as born humans. The 'non-implantable entity' is regarded by Hurlbut as a non-viable artefact, but many of his colleagues on the President's Council on Bioethics disagree⁸.

So the crux of the question is when life begins, a debate that cannot be settled by science. In abstract, this would seem to be the realm of philosophy, but if such debates result in moratoria or bans on research, the medical advances that would surely come from such work will be held in abeyance, and patients with a narrow window of opportunity for treatment will be lost. Their lives are the point. Although the efforts cited here^{1,2} will be criticized as a diversion of good science by politics, I believe all of these attempts to advance and translate medical science should be pursued in parallel.

Irving L. Weissman is at the Stanford Institute

of Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, B257 Beckman Center, 270 Campus Drive, Stanford, California 94305, USA.

e-mail: irv@stanford.edu

1. Chung, Y. *et al.* *Nature* **439**, 216–219 (2006).
2. Meissner, A. & Jaenisch, R. *Nature* **439**, 212–215 (2006).
3. Weissman, I. L. *N. Engl. J. Med.* **346**, 1576–1579 (2002).
4. Hwang, W. S. *et al.* *Science* **308**, 1777–1783 (2005).
5. Committee on Science, Engineering, and Public Policy *Scientific and Medical Aspects of Human Reproductive Cloning* (National Academies Press, Washington DC, 2002).
6. Hurlbut, W. *Monitoring Stem Cell Research* 53–108 (President's Council on Bioethics, Washington DC, 2004).
7. Strumpf, D. *et al.* *Development* **132**, 2093–2102 (2005).
8. Hurlbut, W. *Monitoring Stem Cell Research* 90–93 (President's Council on Bioethics, Washington DC, 2004).

GLOBAL CHANGE

A green source of surprise

David C. Lowe

Living terrestrial vegetation emits large amounts of methane into the atmosphere. This unexpected finding, if confirmed, will have an impact on both greenhouse-gas accounting and research into sources of methane.

On page 187 of this issue, Keppler *et al.*¹ report the remarkable discovery that terrestrial plants emit methane into the atmosphere. Their results are startling, for two reasons. First, because the methane emissions they document occur under normal physiological conditions, in the presence of oxygen, rather than through bacterial action in anoxic environments. Second, because the estimated emissions are large, constituting 10–30% of the annual total of methane entering Earth's atmosphere.

In a series of carefully controlled experiments, Keppler and colleagues used gas chromatography and continuous-flow isotope-ratio mass spectrometry to find that methane is emitted from a wide variety of plant species under oxic conditions. Using ¹³C-labelled acetate substrates, they ruled out the possibility that the methane is produced by anoxic microbial activity. Going further, they showed that this vegetative source depends on sunlight and temperature, with emissions approximately doubling for each rise of 10 °C. The details of the methane-production mechanism are not known, but the authors do demonstrate that emissions are related to the quantity of pectin, a cell-bonding agent, that a plant contains.

To estimate the global methane

emissions from vegetation, Keppler *et al.* make two main assumptions: first, that the emission rates they measured are representative values for short-lived biomass; second, that the

emission estimates can be scaled relative to annual net primary productivity, and can distinguish between different types of environment and average daily hours of sunshine, and between differences in the period of vegetation growth. This type of approach, known as a bottom-up calculation, is commonly used to estimate global emissions from various methane sources and is notorious for producing a wide range of estimates (Fig. 1)². Additional constraints are applied to bottom-up estimates using methane isotopic data and inverse modelling techniques, but the errors remain large.

Most methane is lost from the atmosphere by oxidation, and estimates of this process are used in top-down calculations to deduce the amounts thus removed. For the methane budget to be balanced, the two techniques should agree when the atmosphere is near steady state. But this is rarely the case, as shown by Figure 1. The identification of a new source should prompt a re-examination of the global methane budget, and may ultimately help to reconcile the differences between the bottom-up and top-down techniques.

Meanwhile, Keppler and colleagues' finding¹ helps to account for observations from space of inexplicably large plumes of methane above tropical forests³. They may also explain the current puzzling decrease in the global growth rate of atmospheric methane^{4,5}. Deforestation has led to a dramatic reduction in the Earth's tropical forested area (more than 12% between 1990 and 2000)¹. Keppler *et al.* calculate a corresponding decrease in methane emissions from tropical plants of between 6 million and 20 million tonnes over the same period. During that decade, the rate of methane accumulation in the atmosphere slowed by about 20 million tonnes per year,

a		
Identified methane sources	Estimates ⁸	Range of estimates ²
Total wetlands	145	92–237
Rice agriculture	60	40–100
Ruminant animals	93	80–115
Termites	20	20–20
Biomass burning	52	23–55
Energy generation	95	75–110
Landfills	50	35–73
Ocean	10	10–15
Hydrates (marine and terrestrial)	5	5–10
Total identified sources	530	500–600

b		
Identified methane sinks	Estimates	Range of estimates
Tropospheric oxidation	507	450–510
Stratospheric loss	40	40–46
Soils	30	10–44
Total identified sinks	577	460–580

Total sources–sinks	–47	–80 to +140
----------------------------	------------	--------------------

Figure 1 | Methane sources and sinks. Numbers are millions of tonnes of methane per year. **a**, Estimates⁸ and range of estimates² of annual emissions of methane to the atmosphere from identified sources. **b**, Equivalent figures for methane sinks. The estimates from ref. 8 imply that atmospheric methane is decreasing because identified sources are smaller than the sinks, but this is not confirmed by current measurements of methane in the atmosphere⁵. The ranges in the right-hand column show the extreme values compiled by seven different research groups, as well as estimated total sources and sinks assessed by the Intergovernmental Panel on Climate Change². Note that, for statistical reasons, the sum of the individual source and sink ranges is not the same as the estimated total source and sink ranges. The wide divergence in these figures shows just how ill-defined the methane source inventory and budget are. The new source identified by Keppler *et al.*¹ — methane emitted by vegetation in oxic conditions — is estimated to produce between 63 million and 243 million tonnes per year, but potentially may double count sources listed above.

suggesting that tropical deforestation may have contributed to the decrease.

Methane absorbs solar radiation strongly at infrared wavelengths, and is second only to carbon dioxide in its role in producing an enhanced greenhouse effect and warming the Earth. It also affects the way the atmosphere cleans itself of pollutants, and influences ozone depletion through the production of water vapour in the stratosphere. So methane has been the subject of intense scientific and political scrutiny, and is targeted for emissions controls under the Kyoto Protocol on climate change.

The predominant sources of atmospheric methane are biological. The main ones previously recognized were microbial activity in wetlands (including natural swamps and rice paddy fields) and the eructations of ruminant animals. The dramatic upswing in agriculture required to feed the Earth's growing population has led to huge increases in rice culture and livestock farming in the past 250 years. The result has been large rises in methane emissions from both of these sources.

It was thought that methane production in flooded paddy fields was due to microbial activity in the anoxic environment of the paddy soils. In a 'Kyoto world', in which sources and sinks of greenhouse gases are added and subtracted like the columns in an accountant's report, there are claims that new, 'drier' forms of paddy-field irrigation will lead to reduced methane emissions. But a study of rice plants has shown a strong link between the number and size of leaves on the plant and methane emissions⁶: could the rice plants themselves be as significant a source of methane as the flooded paddy fields?

The implications of Keppler and colleagues' work for the Kyoto Protocol include how reforestation and ruminant animals are treated in methane budgets. Under the Kyoto rules, reforestation since 1990 may be used as a CO₂ sink to offset greenhouse-gas emissions from other sources; we now have the spectre that new forests might increase greenhouse warming through methane emissions rather than decrease it by sequestering CO₂. And in certain countries with large numbers of sheep, cattle and other ruminant livestock, methane constitutes a significant fraction of total greenhouse-gas emissions. In such countries — Ireland and New Zealand, for example — ruminant animals graze on pastures that were originally forested. Given the findings of Keppler *et al.*, it is possible that the forests that once occupied pasture may have produced as much methane as ruminants and grasses on the same land.

The new work will also influence studies of the history of Earth's climate. Indications of past climate are often deduced from analyses of the concentration and isotopic composition of greenhouse gases in tiny air bubbles trapped in polar ice cores. Keppler and colleagues' study shows that, in pre-industrial times, the relative contribution of methane to the

atmosphere by direct emissions from plants could have been much larger than it is today. Measurements of isotopic values in methane derived from Antarctic ice cores show a signal between AD 0 and 1200 that is inconsistent with theories of methane budgets being dominated by wetland sources⁷. A pre-industrial atmosphere containing large contributions of methane derived from vegetation can account for the observed isotopic signal. One of the further avenues of research will centre on the role of methane and vegetation in glacial–interglacial transitions.

This paper¹ will undoubtedly unleash controversy, not the least of which will be political. But there are many scientific questions to be addressed. How could such a potentially large methane source have been overlooked? And what kind of mechanism could produce a highly reduced gas such as methane in an oxic environment? There will be a lively scramble

among researchers for the answers to these and other questions.

David C. Lowe is in the Tropospheric Physics and Chemistry Group, National Institute of Water and Atmospheric Research, Private Bag 14901, Wellington, New Zealand.
e-mail: d.lowe@niwa.co.nz

1. Keppler, F., Hamilton, J. T. G., Brass, M. & Röckmann, T. *Nature* **439**, 187–191 (2006).
2. Prather, M. & Ehrlert, D. in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 239–287 (Cambridge Univ. Press, 2001).
3. Frankenberg, C., Meirink, J. F., van Weele, M., Platt, U. & Wagner, T. *Science* **308**, 1010–1014 (2005).
4. Dlugokencky, E. J., Masarie, K. A., Lang, P. M. & Tans, P. P. *Nature* **393**, 447–450 (1998).
5. Dlugokencky, E. J. *et al. Geophys. Res. Lett.* **30**, doi:10.1029/2003GL018126 (2003).
6. Gogoi, N., Baruah, K. K., Gogoi, B. & Gupta, P. K. *Chemosphere* **59**, 1677–1684 (2005).
7. Ferretti, D. *et al. Science* **309**, 1714–1717 (2005).
8. Mikaloff Fletcher, S. E., Tans, P. P., Bruhwiler, L. M., Miller, J. B. & Heimann, M. *Glob. Biogeochem. Cycles* **18**, doi:10.1029/2004GB002223 (2004).

BEHAVIOUR

Smells, brains and hormones

Gordon M. Shepherd

Contrary to the traditional view, the main olfactory pathway can mediate responses to pheromones as well as to common odours. Recent studies show that pheromone-activated hormonal systems extend widely within the brain.

Pheromones are powerful species-specific chemical signals that organize a wide range of the social conduct of animals, such as mating behaviour, social dominance, aggression, and bonding of a mother with her young. A common belief is that in mammals pheromones are detected only by a specialized sensor in the nose known as the vomeronasal organ, and that the main olfactory epithelium, which lines the nasal cavity, is responsible only for sensing common odours (Fig. 1, overleaf). A long line of under-appreciated work has suggested that this view is too restrictive. Three papers^{1–3} apply the *coup de grâce*, indicating that we need to rethink entirely how pheromones control hormonal responses, not only in mammals generally, but in humans in particular.

In mammals, female mating and reproductive behaviour are controlled by a group of neurons in the hypothalamus, the brain's chief hormonal, or endocrine, control centre. These particular neurons secrete 'luteinizing hormone-releasing hormone' (LHRH; also known as gonadotrophin-releasing hormone, or GnRH) into the hypothalamic–pituitary system to control gonadal and steroidal functions^{4,5}. As they report in *Cell*, Yoon *et al.*¹ and Boehm *et al.*² have developed ingenious methods to trace the brain systems that connect to these neurons.

Yoon *et al.*¹ used a fluorescent virus that is transported only backwards across the junctions

(synapses) between neurons; that is, it will follow the path of a neuronal input to its point of origin. They genetically engineered mice to express a factor specifically in the LHRH neurons that would allow the uptake of the virus into only these cells. So, when the virus is injected into the hypothalamus of the mice, its progress can be traced backwards from the LHRH cells over at least two synapses along the circuits related to these neurons.

The cells that first take up the virus are in the expected hypothalamic areas. However, tracing backwards, labelling occurred in many brain regions, indicating a complex system of mainly olfactory regions and somatosensory areas (dealing with sensations in the body). In the olfactory brain regions, fluorescence was found in the olfactory cortex, the main olfactory bulb, and even out into the main olfactory epithelium. Contrary to predictions from the traditional view, none was seen in the 'accessory' pheromone pathway that originates in the vomeronasal organ. Moreover, behavioural experiments showed that chemosensory modulation of activity in LHRH neurons is primarily through the main olfactory pathway. The results extend previous indications (summarized in ref. 3) that the main olfactory pathway triggers generalized mating behaviour, whereas the vomeronasal pathway mediates specific male and female cues.

Boehm *et al.*² took a different approach. They

engineered mice to contain a gene encoding barley lectin (BL), a tracer molecule that is transferred in both directions across synapses. The BL gene was placed next to the regulatory region for the gene encoding LHRH, so that BL was produced in LHRH cells. In the hypothalamus, expression of BL was limited to some 800 LHRH neurons. From there, it was transferred to the 'vomeronasal amygdala', where the accessory pathway ends, implying that the LHRH neurons are connected to neurons associated with vomeronasal input, either receiving signals from and/or feeding back to the accessory pathway. These same neurons were activated by exposure of the animal to pheromones, showing that these neurons are indeed in the vomeronasal pathway to the LHRH system in the hypothalamus.

There were also some BL-positive neurons in the olfactory cortex and amygdala in the main olfactory pathway, and some of these responded to pheromones. Patterns of neuronal activity markers suggested that pheromone signals can be relayed through the olfactory cortex to LHRH neurons, and that LHRH neurons can in turn feed back through the cortex to the accessory pathway to modulate incoming signals. Many additional brain regions contained BL⁺ neurons, and some of these also contained LHRH⁺ fibres. These findings are consistent with previous work^{4,5} showing that the hypothalamic core has an extraordinarily far-reaching system of neural inputs and outputs.

The question of whether the main olfactory or accessory pathway controls this system has been addressed by Mandiyan *et al.*³. The sensory cells in the main pathway use a membrane channel known as Cnga2 to turn the receptor response into the electrical impulse code. So the authors examined male mating behaviour in mutant mice that lack Cnga2. They found that the mice showed deficiencies in each of the steps involved in mating — sniffing, mounting and intromission — compared with normal controls. Because Cnga2 is not expressed in vomeronasal sensory cells, this indicates that the main olfactory pathway has a "broad and essential" role in mating³.

These results are of interest for several reasons. Hormone-secreting cells in the mammalian brain were first identified in the hypothalamus in the 1950s (reviewed in ref. 6). The

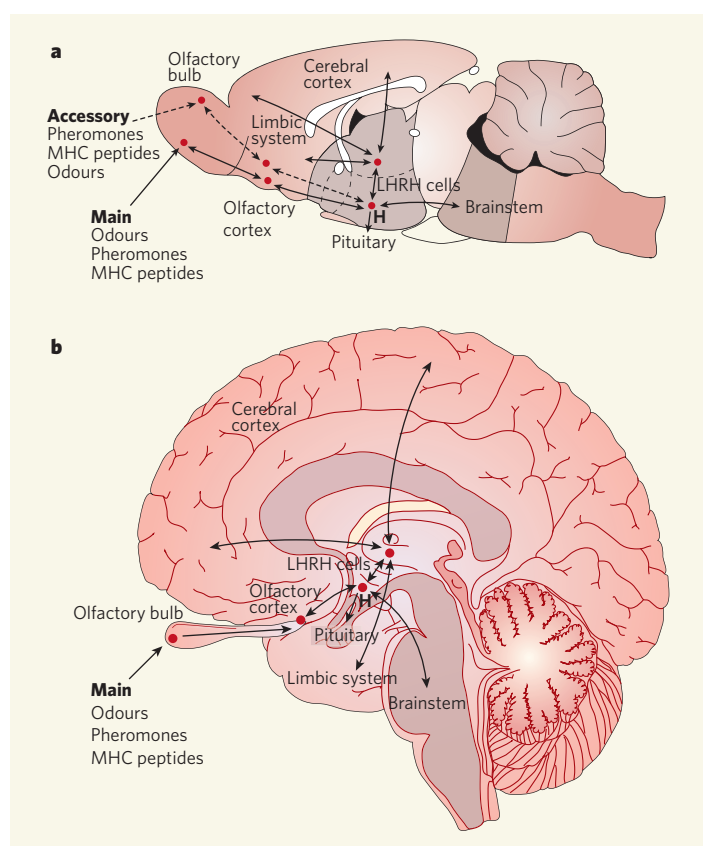


Figure 1 | Brain circuits for pheromones. **a**, A side view of the rodent brain, showing the main (solid lines) and accessory (dashed lines) olfactory systems feeding into the hypothalamus (solid H) and related regions, with connections from the hypothalamus to the cerebrum (for perception and emotion), limbic system and brainstem (for olfactory-guided mating and reproductive behaviour) and pituitary (for coordination of endocrine functions). The new studies^{1–3} highlight the extensive connections of cells expressing luteinizing-hormone releasing hormone (LHRH), showing that the main olfactory pathway is key in mediating pheromonal inputs to them. **b**, Implications for the human brain. In the human brain, only the main olfactory pathway is functional, presumably mediating pheromonal and other olfactory inputs to most of the central LHRH brain systems corresponding to those in the rodent, as well as to cognitive and other regions shown by functional imaging¹⁵.

recent⁴ and current^{1–3} studies show that, in addition to the main and accessory olfactory pathways, this neuroendocrine system includes circuits involved in sexual behaviour, reproduction, arousal, reward, appetite, defensive behaviour and movement. This system for coordinating the neuroendocrine state of the animal with its mating and reproductive activities seems to operate in parallel with the central neurotransmitter systems for noradrenaline, acetylcholine and serotonin that function as a self-governing internal autonomic system for modulating behavioural states throughout the brain. It also interacts intimately with the immune system (see below).

The evidence that the neuroendocrine system can be controlled by both common odours and pheromones acting through the main olfactory pathway confirms and extends previous studies. For example, the suckling pheromone in rats activates a special part of the main olfactory bulb⁷. The male hormone androstenone stimulates mating behaviour in boars despite removal of the vomeronasal

pathway⁸. Urinary pheromones elicit activity patterns in the main olfactory glomerular layer^{9,10}. Mitral cells in the main olfactory pathway respond to urinary social signals¹¹. Pheromones stimulate olfactory as well as vomeronasal sensory neurons¹². And amazingly, a new class of pheromones — non-volatile MHC class I peptides — activate both olfactory and vomeronasal sensory neurons, with the vomeronasal pathway being required for the Bruce effect (pregnancy termination)¹³ and the main pathway being required for mating preference¹⁴.

Together, these findings revolutionize our understanding of the role of smell in controlling the neuroendocrine brain. The traditional distinction that common odours are perceived through the olfactory pathway and pheromones by the vomeronasal pathway is dead. Each pathway must be assessed for a putative pheromone on its own evidence. Humans lack a functional vomeronasal system, but brain scans show that sex pheromones activate wide regions of the human brain, including the cognitive areas¹⁵, implying that, as in the rodent, the pheromone-activated endocrine system is much more extensive than previously realized. We have much more to learn about how intimately neuroendocrine functions, controlled by pheromones acting through our noses, interact with other operations within the brain to control human

behaviour and cognition. ■

Gordon M. Shepherd is in the Department of Neurobiology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA.
e-mail: gordon.shepherd@yale.edu

1. Yoon, H., Enquist, L. W. & Dulac, C. *Cell* **123**, 669–682 (2005).
2. Boehm, U., Zou, Z. & Buck, L. B. *Cell* **123**, 683–695 (2005).
3. Mandiyan, V. S., Coats, J. K. & Shah, N. M. *Nature Neurosci.* **8**, 1660–1662 (2005).
4. Swanson, L. W. *Brain Res.* **886**, 113–164 (2000).
5. Simerly, R. B. *Annu. Rev. Neurosci.* **25**, 507–536 (2002).
6. Scharer, B. *Annu. Rev. Neurosci.* **10**, 1–17 (1987).
7. Teicher, M. H., Stewart, W. B., Kauer, J. S. & Shepherd, G. M. *Brain Res.* **194**, 530–535 (1980).
8. Dorries, K. M., Adkins-Regan, E. & Halpern, B. P. *Brain Behav. Evol.* **49**, 53–62 (1997).
9. Schaefer, M. L., Young, D. A. & Restrepo, D. J. *Neurosci.* **21**, 2481–2487 (2001).
10. Xu, F. *et al. J. Comp. Neurol.* **489**, 491–500 (2005).
11. Lin, Y., Zhang, S. Z., Block, E. & Katz, L. C. *Nature* **434**, 470–477 (2005).
12. Leinders-Zufall, T. *et al. Chem. Senses* **28**, A106 (2003).
13. Kelliher, K. R., Spehr, M., Li, X., Zufall, F. & Leinders-Zufall, T. *Chem. Senses* **30**, A143 (2005).
14. Boehm, T. & Zufall, F. *Trends Neurosci.* doi:10.1016/j.tins.2005.11.006 (2005).
15. Savic, I. *et al. Neuron* **31**, 661–668 (2001).

OBITUARY

Lawrence C. Katz (1956–2005)

Neuroscientist who helped to make sense of sense.

In 1981, just as Larry Katz was beginning his postgraduate career, the Nobel Prize in Physiology or Medicine was awarded to David Hubel and Torsten Wiesel for their pioneering studies of the visual cortex. They had deduced that the vertical columns of neurons in this part of the outer, grey-matter layer of the brain, responsible for processing visual stimuli, are organized into sets — orientation columns and ocular dominance columns — according to their responses to visual stimuli. This discovery came largely through painstaking microelectrode recordings of single neurons, combined with anatomical tracing of connections from eye to brain that revealed a highly organized stripe-like segregation of inputs representing the two eyes in cortical layer 4, the gateway to the cortex.

What was missing from this picture was an understanding of how neuronal connections within the other layers of the cerebral cortex are organized. A cortical column comprises some 10,000 neurons, each making thousands of connections. Aside from hard-won intracellular recording experiments from a handful of single neurons, and beautiful studies that revealed primarily the dendritic morphology of these cortical neurons, little was known about their local axonal branches — the transmitting end of the neuron — let alone about functional synaptic connections. Hubel and Wiesel's work had demonstrated that the columnar organization of the visual cortex could be profoundly perturbed by abnormal visual experience during a critical period of development; but how the cortical columns formed in the first place, or what aspect of altered experience perturbed them, remained mysterious. Answers required new methods of studying the cortical connectivity of several neurons simultaneously, in terms of both anatomy and function.

These were the unresolved questions with which Larry Katz grappled during his tragically short 25-year career, providing many answers and leaving a precious legacy of technological innovation. Katz typified a new breed of scientist, able to combine innovation with a willingness to work on problems, such as those of the wiring of the visual system, that seemed too complex to approach experimentally.

Katz arrived in 1981 at Mark Konishi's laboratory at the California Institute of Technology in Pasadena with a love of neuroethology — the study of brain and behaviour — and with papers on tadpole

and fiddler-crab behaviour, written as an undergraduate at the University of Chicago, under his belt. His intention was to study the brain circuits underlying birdsong. Encouraged by Konishi, however, he switched his focus to the mammalian cerebral cortex, devising in his doctoral thesis of 1984 a striking approach for studying subsets of cortical neurons through the long-distance connections they make with other neurons.

The tracing method that Katz invented used fluorescent latex microspheres that permitted the retrogradely labelled neurons not only to be visualized in slices *in vitro*, but to be simultaneously targeted with a microelectrode and injected with dye. This technique, which Katz perfected as a postdoctoral fellow with Torsten Wiesel at the Rockefeller University, New York, is still popular today, generating images that have graced many journal covers. Quite apart from the beauty of the experiments, the method threw wide a door, opened only a crack by conventional methods, that enabled Katz to examine in exquisite detail the development of horizontal connections between neurons, and show how this development could be influenced by experience.

Not content with an anatomical view of the cortex, in the 1990s Katz — now at his own laboratory at Duke University in Durham, North Carolina — pioneered the application of optical imaging of neurons (using fluorescent calcium or voltage indicators) and photostimulation (using caged glutamate) to probe circuit development. He thus provided new views of cortical circuits, discovering the existence of spontaneously active groups of neurons, known as coactive domains, and probing the emergence of functional horizontal connections within cortical layers 2 and 3.

Katz was never afraid to develop or apply new technology when it was needed for the optimal execution of an experiment. For example, many experiments pointed to a need for correlated neural activity during brain development to fine-tune sets of connections in the cortex that underlie ocular dominance and orientation columns. But *in vivo* evidence for correlated spontaneous firing of ensembles of neurons was missing until Katz perfected a method of multi-electrode recording in conscious, behaving animals.

From 1999 until his untimely death from melanoma on 26 November 2005, Katz's science began to come a full circle, as he returned to his original interest in



neuroethology. In a marvellous conjunction of the best of technology and the best of neuroscience, Katz marshalled his impressive armamentarium to attack the problem of how olfactory signals are processed in the rodent brain. He used the technique of intrinsic signal imaging to reveal a spatial map of odours among structures in the olfactory bulb known as glomeruli, where the first stage of information processing in olfaction occurs. And, in a 2003 tour de force with collaborators Minmin Luo and Michale Fee, he achieved the extraordinary feat of recording neuronal activity in the accessory olfactory bulb of awake, behaving mice. Studying this pathway, which is parallel to the main olfactory pathway and functions as the 'sexual nose', enabled Katz to reveal for the first time the encoding of responses to pheromones — chemical signals transmitted between members of the same species.

Katz's choices of scientific problems centred throughout his career on the key senses of hearing and language (birdsong), sight and, most recently, smell. He appreciated not only on a scientific level the extreme elegance of the brain circuits that process sensory information, but also on a human level their role in connecting us to each other and permitting us to enjoy the world around us to the full. Larry's joy for living and aesthetics no doubt accounted for his love of the sleek design of the Alpha Romeo cars that he drove for years, and of the speed and soaring freedom of flying an aeroplane. He was also an avid fisherman, and we can imagine him even today with a prize albacore or up to his knees in a rushing stream high up in the wilderness, reeling in a trophy trout.

Carla J. Shatz

Carla J. Shatz is in the Department of Neurobiology, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA.

e-mail: carla_shatz@hms.harvard.edu

B. ARENKIEL

BRIEF COMMUNICATIONS

Teaching in tandem-running ants

Tapping into the dialogue between leader and follower reveals an unexpected social skill.

The ant *Temnothorax albipennis* uses a technique known as tandem running to lead another ant from the nest to food — with signals between the two ants controlling both the speed and course of the run. Here we analyse the results of this communication and show that tandem running is an example of teaching, to our knowledge the first in a non-human animal^{1–3}, that involves bidirectional feedback between teacher and pupil. This behaviour indicates that it could be the value of information, rather than the constraint of brain size, that has influenced the evolution of teaching.

An individual is a teacher if it modifies its behaviour in the presence of a naïve observer, at some initial cost to itself, in order to set an example so that the other individual can learn more quickly¹. We suggest that teaching also involves bidirectional feedback between teacher and pupil. To test whether tandem running fulfils these criteria, we measured the acceleration of leaders and followers in response to the stimuli they present to one another.

In our experiments, tandem leaders knew the location of food but tandem followers were naïve (for details, see supplementary information). We found that the leader only continued the tandem run when frequently tapped on her legs and abdomen (gaster) by the following ant's antennae^{4,5} (Fig. 1a, inset; for movie, see supplementary information). The tandem leader therefore modifies its behaviour in the presence of the follower.

We also found that tandem leading imposes a cost on the leader in that she can proceed four times faster from the nest to the food when not encumbered by a follower (median speed was 1.8 mm s^{-1} , compared with 8.4 mm s^{-1} ; Mann–Whitney test, $P < 0.0001$). Tandems are slowed by frequent pauses in which the follower loops around, probably in a search for landmarks^{6,7}; followers make larger looping movements than do individual explorers (Mann–Whitney test, $P < 0.001$; see supplementary information). Indeed, the speed of leader and follower between pauses was significantly higher in the presence of conspicuous landmarks than the speed of controls in the absence of landmarks (mean speed was 7.6 mm s^{-1} , compared with 5.8 mm s^{-1} ; analysis of variance (ANOVA), d.f. = 1, $F = 22.2$, $P < 0.0001$).

We next tested whether the leader provides a demonstration of how to find food. Followers found food more quickly when tandem

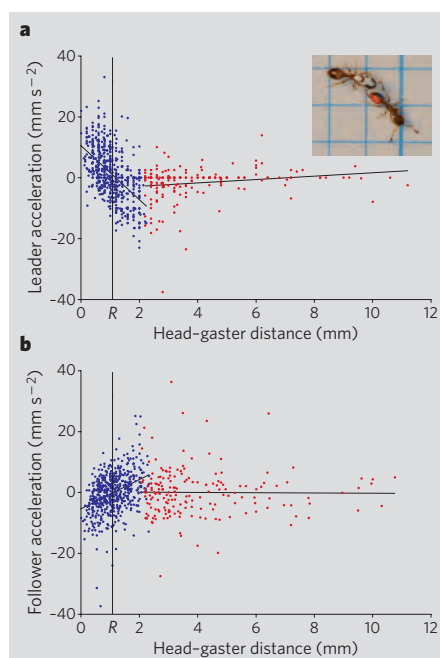


Figure 1 | Tandem running in the ant *Temnothorax albipennis*. **a, b,** Acceleration by ten ants (results pooled) that were **a,** leaders, or **b,** followers, in response to mutual stimuli, as a function of the separation between the leader's gaster and the follower's head. When within twice the distance of the antennal reach (R ; average R , 1 mm) of the follower (blue circles), the ants are too close so the follower decelerates and the leader accelerates; when further apart, the follower accelerates and the leader decelerates. At the average reach of the follower's antennae, both move at constant velocity; when separated by more than twice the antennal reach of followers (red circles), leaders stop and wait as followers loop around. Inset, tandem run over 2-mm squares (for video, see supplementary information).

running than when searching alone (mean time was 201 s, compared with 310 s; ANOVA, d.f. = 1, $F = 6.40$, $P = 0.016$), indicating that the follower learns more quickly as a result of the leader's help. Moreover, the follower's return path to the nest does not replicate the route of the tandem run: its journey is generally faster and more linear (in 6 out of 8 cases) than that of its leader before the tandem run (see supplementary information). The follower therefore learns the food's location sooner and gains general knowledge about the surroundings of the nest as a result of the leader's example.

Bidirectional feedback between the leader

and follower is evident from their patterns of acceleration and deceleration as a function of the strength of the stimuli they present to one another (Fig. 1): when the gap between the leader and follower grows too large, the former decelerates and the latter accelerates; both move at the same speed when at the maximal antennal reach of the follower.

Together, our results show that the leader's performance fulfils all the criteria for teaching, with the follower acting as pupil. The lessons learned by tandem followers are transferred when they become tandem leaders⁸, so — although tandem runs are slow — they propagate time-saving knowledge among foragers.

Temnothorax albipennis workers may also carry nestmates⁹. Such pairs move three times faster than tandems, but carried ants do not teach others — perhaps because they cannot learn while their head is inverted and pointing backwards during carrying. Other species of ant workers that do not use tandem running⁴ are carried with their heads upright: these can learn a route and later recruit others¹⁰.

Bidirectional feedback between teacher and pupil distinguishes teaching from broadcasting. Most recruitment in large ant societies is broadcasting (for example, through pheromone trails)⁴, which is effective in big groups. But in small societies, where information is valuable and easily lost, teaching works better. Our identification of teaching behaviour in an ant shows that a big brain is not a prerequisite.

Nigel R. Franks, Tom Richardson

School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK

e-mail: nigel.franks@bristol.ac.uk

- Caro, T. M. & Hauser, M. D. Q. *Rev. Biol.* **67**, 151–174 (1992).
- Shettleworth, S. J. *Cognition, Evolution and Behaviour* 475–522 (Oxford Univ. Press, 1998).
- Galef, B. G. Jr, Whiskin, E. E. & Dewar, G. *Anim. Behav.* **70**, 91–96 (2005).
- Hölldobler, B. & Wilson, E. O. *The Ants* (Harvard Univ. Press, Cambridge, Massachusetts, 1990).
- Möglich, M. & Maschwitz, U. & Hölldobler, B. *Science* **186**, 1046–1047 (1974).
- Pratt, S. C., Brooks, S. E. & Franks, N. R. *Ethology* **107**, 1125–1136 (2001).
- McLellan, M. A., Pratt, S. C. & Franks, N. R. *Insectes Soc.* **49**, 203–208 (2002).
- Möglich, M. *Insectes Soc.* **25**, 205–225 (1978).
- Pratt, S. C., Mallon, E. B., Sumpter, D. J. T. & Franks, N. R. *Behav. Ecol. Sociobiol.* **52**, 117–127 (2002).
- Möglich, M. & Hölldobler, B. *Psyche* **81**, 219–236 (1974).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/439153a

Hit-and-run planetary collisions

Erik Asphaug¹, Craig B. Agnor¹ & Quentin Williams¹

Terrestrial planet formation is believed to have concluded in our Solar System with about 10 million to 100 million years of giant impacts, where hundreds of Moon- to Mars-sized planetary embryos acquired random velocities through gravitational encounters and resonances with one another and with Jupiter. This led to planet-crossing orbits and collisions that produced the four terrestrial planets, the Moon and asteroids. But here we show that colliding planets do not simply merge, as is commonly assumed. In many cases, the smaller planet escapes from the collision highly deformed, spun up, depressurized from equilibrium, stripped of its outer layers, and sometimes pulled apart into a chain of diverse objects. Remnants of these 'hit-and-run' collisions are predicted to be common among remnant planet-forming populations, and thus to be relevant to asteroid formation and meteorite petrogenesis.

Recent modelling¹ shows that for the random velocities expected^{2,3} following the initial runaway-oligarchic stage of terrestrial planet growth^{4–6}, about half of giant impacts result in no net mass accumulation. In many cases, especially at grazing incidence, the impactor emerges disrupted (for example, ref. 7) but unaccreted. We focus on two principal consequences of impactor disruption: (1) unloading of a surviving planet-sized impactor from hydrostatic pressure over the encounter timescale (hours), which leads to fracture, igneous alteration, and possible melting and petrogenesis; and (2) density-driven material segregation, which occurs when impactors lose their outer layers owing to tides, shocks and shears.

From Roche encounters to hypervelocity collisions

Giant impacts, defined as collisions between planets of comparable size, range from grazing events (impact angle $\xi \approx 90^\circ$; see Fig. 1), where gravitational stresses dominate, to direct hits ($\xi \approx 0^\circ$), where shocks dominate. The best-studied giant impact of all, the proposed origin of the Moon (reviewed in refs 8 and 9), appears to require a collision of intermediate ξ and an impact velocity v_{imp} just barely above the two-body escape velocity $v_{\text{esc}} = \sqrt{2G(M+m)/(R+r)}$, where M and m are the masses of the colliding planets, G is the gravitational constant, and R and r are their radii. In one scenario¹⁰, a Mars-sized planet (about half an Earth diameter, or about one-tenth an Earth mass, $M_\oplus$) collides with the proto-Earth at about $\xi \approx 45^\circ$ and $v_{\text{imp}} \approx 1.05v_{\text{esc}}$ and is sheared by gravitational and mechanical torques and shocked to high pressure. A protolunar disk consisting about equally of impactor mantle and target mantle is ejected, while the remainder (including nearly all the impactor's core) accretes. This segregation of rock from iron, which accounts in this scenario for the iron-poor Moon, is also observed in models of collisions that are non-accretionary, leading us to explore generally how giant impacts process and segregate unaccreted materials.

The end-member archetypes of planetary collisions are hypervelocity impacts and disruptive tidal encounters inside the Roche limit. Both end members were exhibited by comet Shoemaker–Levy 9 (S–L/9). Much smaller than Jupiter and half as dense (interloper radius $r \approx 1$ km and density $\rho_i \approx 0.6 \text{ g cm}^{-3}$; ref. 11), S–L/9 fragmented deep inside the Roche limit at periaapse ($a = 1.31R \approx 0.4 R_{\text{Roche}}$) into ~ 20 gravitationally bound clumps, on 8 July 1992. Here $R_{\text{Roche}} = 2.46R(\rho_t/\rho_i)^{1/3}$ is the Roche limit for tidal disruption, R is the target radius (in this case Jupiter), and ρ_t and ρ_i are the target and impactor densities.

As luck would have it, after a two-year eccentric orbit about Jupiter S–L/9 became the one observed example of a planet-scale hypervelocity collision. The fragments of tidal disruption, now distributed as a 'string of pearls', came back and struck Jupiter at $\sim 60 \text{ km s}^{-1}$, producing a series of large dark blemishes observable with backyard telescopes. The tidal break-up involved gravitational unloading over the course of hours and the formation of a distinctive family of comets; the impacts involved intense shock compression in a tenth of a second and a few billion tons of shock-evolved compounds added to Jupiter. For small interlopers such as comets, planetary encounters are generally 'hit or miss'. For larger impactors ($r \lesssim R$; see Fig. 1), on the other hand, there is a broad range of intermediate outcomes we now explore.

Larger impactors

Tidal disruption can only happen to interlopers larger than a given size, for a given strength¹², so that if small comets commonly suffer tidal disruption¹³ then one expects the process to be a common aspect of planet formation. But in late-stage giant impacts following runaway-oligarchic growth^{4–6}, impactors and targets are of comparable size, and cannot get deep inside the tidal field without colliding. It has therefore been argued that massive impactors escape tidal disruption¹⁴. However, massive impactors are observed in the calculations presented below to suffer profound tidal effects during giant impacts, in addition to profound shocks and mechanical shears—more so than their targets, which are notoriously difficult to disrupt by collision¹⁴.

Massive impactors are much more easily disrupted than the targets they strike, for three reasons. (1) The ratio of tidal stress to self-gravitational stress between colliding bodies of masses m and M , of similar composition, varies inversely with mass, so the impactor bears the brunt of any tidal deformation. (2) Global shock intensities are concentrated in the impactor, approximately in proportion to its smaller volume, because shocks propagate about equally in both directions from the contact plane. (3) Strong mechanical shears, which arise because part of a giant impactor simply 'misses' the target (see Fig. 1), lead to differential stresses that are inversely proportional to the decelerated mass. Overall, an impactor half the size of a target will suffer disruptive consequences almost an order of magnitude more severe. Below we show case studies of typical planetary embryos colliding but not accreting: in many cases, the impactor becomes a novel planet or a family of novel planets, plus copious debris.

¹Earth Sciences Department, Institute for Geophysics and Planetary Physics, 1156 High St, University of California, Santa Cruz, California 95064, USA.

Unloading

Planets possess great specific enthalpy in their compressed interiors. When a planet is disrupted, it unloads from gravitational equilibrium, which liberates significant quantities of internal energy separate from the kinetic energy of impact. For example, the ~ 20 fragments of S-L/9 exhibited prolific and energetic dust production¹⁵, probably owing to the sudden exposure of volatile-rich interior surfaces released from an overburden pressure $P \approx 0.001$ bar. A planetary embryo, if disrupted by a similar process, would unload from pressures millions of times greater, and if rich in volatiles¹⁶ might erupt. The energy per unit mass of the product of decompression¹⁷ is $\int dP/\rho$, which for constant density (up to the onset of vaporization) is $\sim P/\rho \approx G\rho r^2$. Material unloaded from the base of a Mars-sized planet's mantle releases $\sim 2 \times 10^{10}$ erg g⁻¹, about the energy density of TNT.

Unlike shocks, which deposit energy locally, this unloading energy is available throughout the interior, leading to thermophysical events of different scale and character to those associated with impact shocks. The energy of unloading is proportional to the square of a disrupted planet's radius; unloading of a Neptune-mass planet's mantle would release over 10^{11} erg g⁻¹. In stellar encounters collisional unloading is particularly dramatic, for consideration regarding turn-of-the-century "planetesimal" concepts¹⁸ of planet formation.

To evaluate the significance of pressure unloading in terrestrial collisions, we first simulate a purely tidal encounter with no shock pressures. We use the smooth particle hydrodynamics (SPH) code of Benz¹⁹ to model a differentiated Moon-sized planet (70 wt% rock mantle, 30 wt% Fe core) suffering a close (non-impacting) encounter with a Mars-sized planet (Fig. 2a). The method and equation of state are those of ref. 10. Figure 2b plots the evolution of the interior pressure over the course of the encounter; the interloper experiences

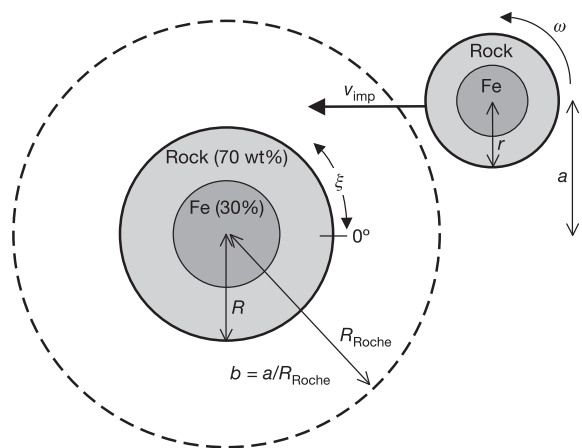


Figure 1 | Planetary embryos of comparable diameter are believed to have collided in giant impacts in the late stage of Solar System formation. This occurs following runaway-oligarchic growth^{4–6}. We model colliding planets as differentiated, equilibrated bodies composed of 30 wt% iron cores surrounded by 70 wt% rocky mantles. Here R_{Roche} is the Roche limit (see text), ω is the rotation rate of the impactor (assumed zero in the calculations presented below), a is the closest approach of the encountering centres of masses (focusing included), v_{imp} is the impact velocity ($v_{\text{imp}}^2 \approx v_{\infty}^2 + v_{\text{esc}}^2$ where v_{∞} is the velocity at infinity, that is, the random velocity of the encountering planets and $v_{\text{esc}} = \sqrt{2G(M+m)/(R+r)}$ is the two-body mutual escape velocity), M and m are the masses of the planets, R and r are their radii, and ξ is the angle of first contact for undeformable homogeneous spheres, where 90° is grazing, 0° is head-on. Although the usual rule applies that $\xi = 45^\circ$ is the most probable, there is an important scale difference relative to better-studied cratering events. When $r \approx R$, half of a 45° impactor 'misses' the planet, whereas when $r \ll R$ all of a 45° impactor collides (a typical cratering event). Planets are gravitationally compressed before our simulations, so that bulk density ρ is slightly greater for the planet with the larger mass, M .

a transient global pressure drop of ~ 30 – 50% , from tides alone, for about an hour. Internal pressure recovers to only $\sim 80\%$ of its pre-encounter value because of spin-up and removal of the outer layers. Pressure release melting at a global scale might ensue if the

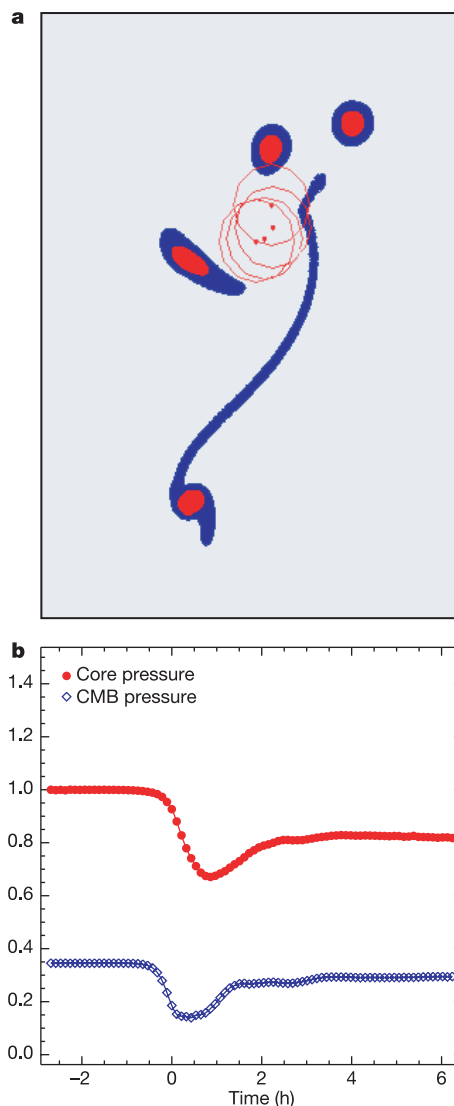


Figure 2 | A Moon-sized differentiated planet ($M = 0.01M_{\oplus}$) grazing a Mars-sized ($0.1M_{\oplus}$) planet, resulting in mass loss, spin-up and global pressure unloading. Here $M_{\oplus}$ is the mass of the Earth. There is no physical contact, only gravitation and unloading. The interloper is a non-rotating, isostatically and thermally equilibrated sphere. Because the interloper does not impact, the central body is modelled as a sphere in this calculation. The closest-approach velocity is $1.05v_{\text{esc}}$ and the closest approach distance is $a = 1.05(R+r)$. **a**, The two planets are shown in the centre-of-mass frame, with the larger planet being displaced towards the top of the figure while the smaller planet swings by towards the bottom and is severely deformed. The larger planet is shown as an open red circle, with its centre of mass shown by a red dot; the smaller planet is shown in blue and red, where blue is the mantle rock and red is the iron core, as revealed in this slice through the symmetry plane of the collision. This panel shows a series of 'snapshots' of the event. Deformation is, in this case, largely a consequence of tidal stress, and thus low-density materials in the outer layers are deformed more severely than the denser interior and are partially removed. **b**, Interior pressures versus time are plotted, averaged over the core–mantle boundary (CMB) region (blue line) and over the central core region (red line) of the impactor. Here P is the pressure in each region of the interloper, and P_c is the initial hydrostatic central pressure. The pressure unloads throughout the interloper, by 30–50% for about an hour during periapse, and permanently by about 20% as a result of mass loss and spin-up.

planet is already partially molten. In the removed materials, unloading is permanent and almost total, and equally abrupt—a scenario for phreatic igneous processes discussed below.

For closer, impacting encounters, gravitational stresses increase, and shock and mechanical stresses also become intense. We model impacts using the same SPH technique, starting with two differentiated, equilibrated planets of identical composition and comparable size. Characteristic outcomes (see refs 1, 20) of such collisions are shown in Fig. 3; both involve a Mars-sized target struck at $\xi = 30^\circ$ at typical velocities. In the first scenario (Fig. 3a), the impactor is one-half the target mass and $v_{\text{imp}} = 1.5v_{\text{esc}}$ (that is, the random relative velocity $v_\infty \approx 1.1v_{\text{esc}}$). It emerges more-or-less intact but stripped of its outer layers and spun into fast rotation. Figure 3b shows a faster impactor ($v_{\text{imp}} = 2v_{\text{esc}}$; $v_\infty \approx 1.7v_{\text{esc}}$) one-tenth the target mass which disrupts into a fractionated S–L/9-like chain of fragments, with iron-rich bodies at the centre (shown red) and mantle-rich bodies in the ‘wings’. There is no net accretion in either collision; iron fraction is increased in both planets due to mantle loss, most notably from the impactor. A report on the detailed outcomes of hundreds of representative collisions is forthcoming.

Degassing

When impactors and their removed materials depressurize from pre-impact equilibrium conditions, they might degas. Figure 4 shows equilibrium equations of state computed for molten early mantle containing 1, 5 and 10 wt% water—plausible conditions for young planetary embryos. The limited solubility of water in silicate liquids at low pressures means that the mantle of a hot primitive embryo is probably degassed at shallow levels, whereas the deep mantle can retain abundant water. Planetary embryos suffering severe mass loss, spin-up or fragmentation might, according to Fig. 2b, experience a global hydrothermal event in the hours during and following depressurization. According to Fig. 4, a core-mantle pressure drop

of 50% in a water-bearing Moon-sized object would initiate extensive deep-mantle degassing. As for materials no longer bound to an impactor (Figs 2a and 3), there is little if any post-encounter pressure support. Removed materials would be almost totally unloaded from their equilibrium pressure, and if molten would efficiently degas throughout—a recipe for dry, igneous, fine-scale meteoritic debris.

Stress accommodation

Unloading from equilibrium pressure requires deformation over the encounter timescale, and there exist conflicting analyses of whether this is possible. SPH simulations similar to those above were used²¹ to argue that bulk viscosity prevents tidal disruption, but it was later argued²² that the use of numerical damping in place of rheological viscosity in those simulations led to artificial resistance to deformation, and that tidal disruption does occur. SPH is not the best tool for studying viscous deformation, so we address this topic with a straightforward geophysical analysis of a viscoelastic impactor succumbing to deformational stresses during the course of a hit-and-run collision. To be conservative, we consider the tidal stresses, apart from shocks and mechanical shears, that can disrupt an impactor.

For incompressible inviscid impactors much smaller than the target, tidal-induced gravitational deformation is invariant with size when distance is normalized to R_{Roche} and time normalized to the gravitational timescale τ_{grav} , or equivalently v_∞ to v_{esc} (ref. 11). Assuming $v_\infty \propto v_{\text{esc}}$, the encounter timescale is $\tau_{\text{grav}} = (G\rho)^{-1/2}$. The invariance arises because tidal stress scales with $G\rho^2 b^{-3} r^2$, where b is the distance to the target centre-of-mass normalized to R_{Roche} , while the overburden stress that is acted against scales with $G\rho^2 r^2$. Real planets also resist deformation through strength and viscosity, in addition to self gravity, but these forces, we now show, are small compared with unloading stresses at the deformation rates expected during late-stage collisions.

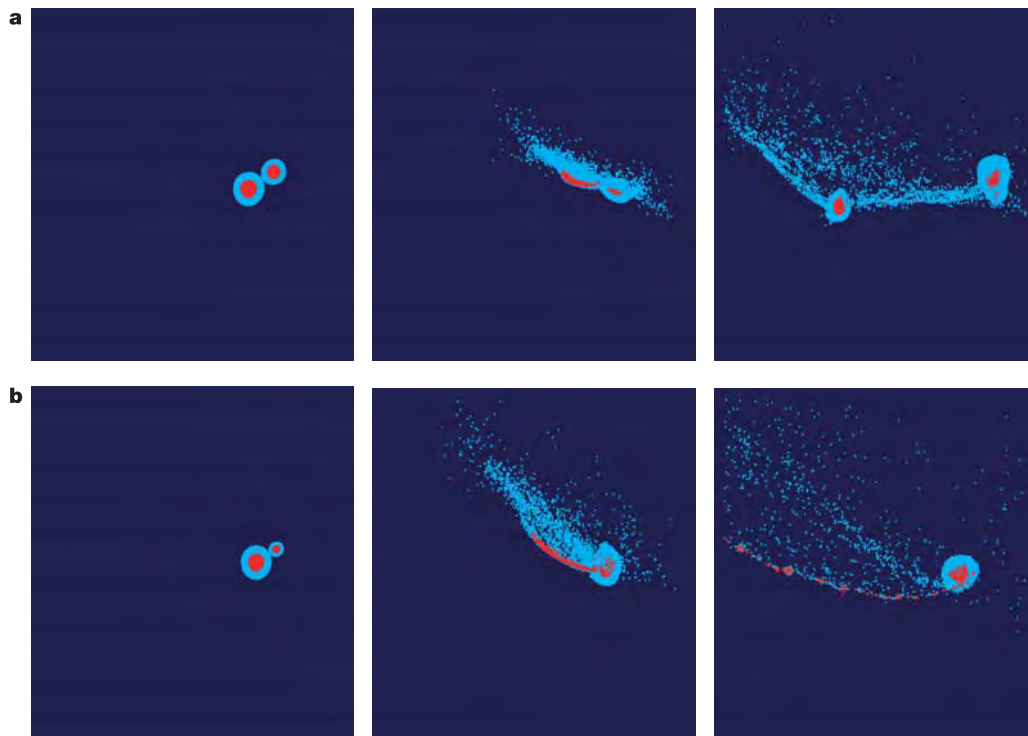


Figure 3 | Two typical collisions involving differentiated planetary embryos. The target mass is $M = 0.10M_{\oplus} = 5.98 \times 10^{26}$ g and the impactor mass is **a**, $m = M/2$, and **b**, $m = M/10$. Blue is mantle rock and red is core iron, this time showing all SPH particles overplotted. Impact velocities are typical of terrestrial planet formation, **a**, $v = 1.5v_{\text{esc}}$ and **b**,

$v = 2.0v_{\text{esc}}$. Impact angle is $\xi = 30^\circ$ in both cases. The simulations are shown before, during, and 3 h after the impact, in side view. Mantle (blue) is lost in some degree from all major remnants of non-accretionary collisions; in **b**, a chain of iron-enriched bodies derive from the impactor.

Elastic failure

Let $\eta = \eta_{\text{eff}}$ be the effective newtonian viscosity at given stress σ_{ij} and temperature T , and E the elastic modulus. If the visco-elastic Maxwell time $\tau_{\text{Maxwell}} = \eta/E > \tau_{\text{grav}}$, the response is elastic and disruption requires brittle fragmentation. For $E \approx 10^{10} \text{ dyn cm}^{-2}$ and τ_{grav} of several hours, elastic response applies for $\eta > 10^{14}$ poise (1 poise = $1 \text{ g cm}^{-1} \text{ s}^{-1}$), the case we consider first. For a given encounter distance a , tidal stress increases with r^2 , where r is the impactor radius; this led to Jeffreys' result¹² that elastic rocks larger than $\sim 200 \text{ km}$ fracture during close encounters with Earth.

Interestingly, the brittle unloading of the impactor's equilibrium overburden stress over τ_{grav} can itself lead to dynamic fragmentation, apart from shock damage. If an elastic mantle obeys Weibull failure statistics, then the number of active flaws per unit volume is $n(\epsilon) = k\epsilon^m$, where $\epsilon = \sigma/E$ and σ is the flaw activation stress. If the characteristic mantle stress $\sigma \approx G\rho^2 r^2$ unloads over τ_{grav} , then the characteristic strain rate is

$$\dot{\epsilon} = G^{3/2} \rho^{5/2} r^2 / E \quad (1)$$

which gives²³ a characteristic fragment size $L \approx 6c_g \alpha^{-1/(m+3)} \dot{\epsilon}^{-m/(m+3)} / (m+2)$, where $\alpha = 8\pi c_g^3 k / [(m+1)(m+2)(m+3)]$ and c_g is the crack growth velocity, about half the sound speed. Fragment size thus decreases with almost the square of disrupted planet size, as $m \geq 6$ for most geologic materials²⁴.

Fragmentation also results in response to shear strains whenever part of an impactor is decelerated while another part continues on relatively unimpeded. This shear strain rate can be approximated as the decelerated impact velocity divided by the projectile radius, or $\dot{\epsilon} = v_{\text{imp}}/r$, in which case fragment size decreases approximately linearly with r for a given impact speed. These shear strain rates are in most scenarios much larger than those resulting from stress unloading, and can in principle result in even smaller fragment sizes. But shear strain may be localized, whereas equilibrium unloading is rather uniformly distributed, so unloading establishes the

upper limit on expected fragment size. A cold 500-km-diameter basalt sphere ($k = 4 \times 10^{29} \text{ cm}^{-3}$, $m = 9$) cracks into $\sim 200\text{-m}$ fragments if unloaded of its equilibrium stress over τ_{grav} , and a 1,000-km sphere cracks into $\sim 70\text{-m}$ fragments. Weibull coefficients for other rocks give comparable sizes. This indicates that no large monoliths survive following the disruption of large solid bodies, even without considering shock damage. Shocks, shears and degassing stresses further comminute unloaded rocky materials. Instant rubble piles form if these new fragments do not disperse, and families of small asteroids form otherwise. This is of relevance to the 'missing mantle' paradox (discussed further below), where mantle material is missing from the meteorite collection and not readily identified among asteroids. One offered explanation²⁵ is that the excavated mantle rock is fragile and therefore ground down to small sizes by a subsequent collisional cascade—something made easier if fragmentation is a natural outcome of mantle unloading.

Viscous failure

During late-stage accretion, impact heating and short-half-life radionuclide decay kept embryos hot, perhaps molten, for millions of years; viscous deformation is therefore a more likely scenario for large-scale events. Viscosity decreases with $e^{-T_0/T}$, where T is temperature and T_0 a reference temperature. The effects of partial melt and water content markedly increase the strain rates at which silicates deform, and while grain sizes and melt distribution have complex rheologic effects, characteristic shifts in strain rates for even 10% partial melting are well over an order of magnitude under hydrous conditions²⁶. For power-law fluids $\dot{\epsilon} \propto \sigma^n$, and the effective viscosity $\eta_{\text{eff}} \propto \sigma^{1-n}$ decreases with about the square of the applied stress ($n \approx 3$ for cold ice and dry quartzite). Effective viscosity is thus low under high differential stress, especially the conditions of planetary unloading.

To disrupt, deformation strain ϵ_{def} must accrue over a few times τ_{grav} . As an order of magnitude estimate, let $\epsilon_{\text{def}} \approx 10$ for disruption. The maximum viscosity η_{lim} accommodating this deformation is approximately the stress that must be unloaded, $\sigma \approx G\rho^2 r^2$, divided by the required strain rate, $\dot{\epsilon} \approx \epsilon_{\text{def}}/\tau_{\text{grav}}$. The following viscosity thus limits tidal disruption:

$$\eta_{\text{lim}} \approx \epsilon_{\text{def}}^{-1} \sqrt{G\rho^3} r^2 \quad (2)$$

Viscous strains $\epsilon_{\text{def}} > 10$ accrue when $\eta < \eta_{\text{lim}} \approx 10^{12} (r/1,000 \text{ km})^2$ poise ($\text{g cm}^{-1} \text{ s}^{-1}$). As benchmarks, models of early-Earth convection assume mid-mantle viscosities $\sim 10^9$ poise (ref. 27), and $\eta \approx 10^9\text{--}10^{13}$ poise is used²⁸ to model Io's present asthenosphere.

Analyses in viscous and brittle regimes therefore suggest that embryos the size of the largest present asteroids were vulnerable to disruptive deformation from tides alone, let alone impact shocks and disruptive shears, as long as larger 'target' embryos were around. Hit-and-run planetary collisions comparable in scale to the proposed Moon-forming event^{8–10} may well have happened on the way to terrestrial planet formation, as they are statistically common for expected impact parameters¹. The most recognizable signatures of these disrupted and fragmented impactors would then be found primarily among populations where unaccreted relics survive.

Giant impacts in the asteroid belt

The original asteroid belt is believed to have accreted to form bodies with diameters of $\sim 1,000 \text{ km}$ (or larger²⁹) and then to have dynamically lost most of its mass, thereby 'fossilizing' an early stage of accumulation³⁰. Because accretion in the late stage is favoured¹ at low ξ , the 'winners' of accretion (finished planets) tend to consist of shock-processed materials acquired by direct hits, whereas 'losers'—for example, post-accretion asteroid populations—become relatively enriched in unaccreted collisional material forming when bodies of similar size collide. Meteorites that are derived from large, differentiated asteroids are thus expected to contain a mix of relatively unshocked materials worked by gravitational unloading

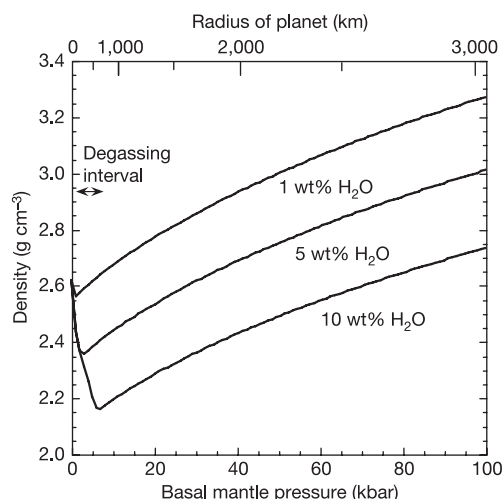


Figure 4 | The pressures at which degassing initiates. These are evaluated for variable water contents, and in terms of the radius of planetesimals containing 30 wt% Fe cores. We calculate the density of silicate melt with $(\text{Mg}_{0.9}\text{Fe}_{0.1})\text{SiO}_3$ -composition which is initially at the core–mantle boundary (CMB) pressure. The initial Moon-sized impacting embryo used in the hit-and-run simulations of Figs 2 and 3b plots at $R = 1,500 \text{ km}$. The initial embryo is assumed to be largely undegassed at depth, with water lowering the bulk density relative to the $(\text{Mg}_{0.9}\text{Fe}_{0.1})\text{SiO}_3$ endmember; when mantle is unloaded to below a given value (~ 6 , ~ 4 and $\sim 2 \text{ kbar}$ for 10, 5 and 1 wt% H_2O respectively), the melt becomes oversaturated with water and will begin to degas, increasing the density of the residual melt. The timescale of the pressure unloading (Fig. 2) suggests that degassing will be runaway rather than transitory, as the timescale for bubble nucleation is of the order of 100 s. Density, elasticity and water solubility data are from refs 41–43.

and mechanical shearing, together with the shocked remnants from direct collisions. As noted above, the small, undifferentiated members of the original asteroid population will not commonly undergo hit-and-run collisions, simply because when $r \ll R$ encounters are either purely tidal, and gentle for these size scales, or else are hypervelocity cratering events (for example, S–L/9).

Disruption, pressure release and fractionation observed in our simulations, and associated hydrothermal activity and possible degassing, can be related to an array of quandaries^{25,31–34} regarding meteorites deriving from large parent bodies. One straightforward puzzle is the apparent absence of mantle-derived meteorites ($\leq 1\%$ of non-chondritic falls) when so many irons and stony-irons have been excavated from core regions ($\geq 10\%$ of non-chondritic falls). The major volume of these disrupted planets appears absent. This is answered in two ways by the models presented above. First, liberated mantle, if brittle, is fragmented by unloading to relatively small sizes (equation (1)). If hot and viscous (equation (2)), liberated mantle rocks are likely to be melted by unloading and hydrothermally altered as water exsolves, and might not be recognizable as mantle rock. Second, the global mechanical shearing suffered by an impactor greatly increases the surface area of the core–mantle boundary during a hit-and-run collision (for example, Fig. 3b), thereby mixing iron and silicate and increasing the abundance and diversity of stony-iron meteorites.

A related puzzle is how to remove entire mantles from some differentiated asteroids while preserving others. No fewer than 10 and as many as 100 distinct iron reservoirs, presumably embryonic cores, are sampled by iron meteorites (ref. 24 and references therein). A number of large M-class asteroids (for example, 216 Kleopatra³⁵ and 16 Psyche) are believed to be metallic, as are most smaller M-class asteroids³⁶. A serious problem arises if impact is used to excavate iron from the cores of so many asteroids, because it becomes statistically impossible for the 530-km-diameter main-belt asteroid 4 Vesta to preserve its signature basaltic crust³⁷ while neighbouring asteroids are stripped bare. Even if Vesta managed to dodge all catastrophic impactors, any impacting population includes a power-law distribution of smaller sizes³⁸, so Vesta must also have dodged myriad subcatastrophic collisions that would scramble, if not disperse, its basaltic crust.

Explaining Vesta's surviving crust, while accommodating widespread asteroid mantle removal, appears intractable. However, if these metallic asteroids and associated remnants began as one or more impactors disrupted by larger targets (for example, the disrupted interloper in Fig. 3b) and not vice versa, then Vesta only had to avoid collisions with larger bodies until any Moon-sized or larger embryos were ejected from the main belt²⁹. The corollary is that meteoroids massive and energetic enough to catastrophically disrupt Vesta-sized asteroids were relatively rare, and that Vesta's bombardment experience (one hemispheric impact basin formed over its history³⁹) is typical rather than unusually lucky.

A third puzzle is the evidence for widespread global melting of asteroids on the basis of oxygen isotope homogenization in meteorite parent bodies⁴⁰. It has been rightly argued³³ that impacts are inefficient at global heating of asteroids, because impact shock can only deliver so much heat before disrupting and dispersing a body of this size, and because shock heating is local. Global melting of asteroids by short-lived radionuclide decay is even more controversial, given the critical timing constraints. Gravitational unloading can melt a body by pressure release throughout its deep interior, a process warranting more detailed examination on the basis of our studies so far.

Discussion

We have revisited planetary giant impacts from the perspective of the impactor, and identify an additional scenario for asteroid formation and planetary evolution where the smaller body is disrupted by the larger. This mode of disruption is complex, and some of the

thermophysical effects (pressure unloading and removal of the outer layers) require more detailed exploration. But owing to the prevalence of 'hit and run' collisions at expected encounter geometries and velocities¹, the products of impactor disruption are likely to be common among unaccreted planetary populations—asteroids, meteorites and perhaps the smallest planets.

Received 13 July; accepted 6 October 2005.

1. Agnor, C. & Asphaug, E. Accretion efficiency during planetary collisions. *Astrophys. J.* **613**, L157–L160 (2004).
2. Wetherill, G. W. Occurrence of giant impacts during the growth of the terrestrial planets. *Science* **228**, 877–879 (1985).
3. Agnor, C. B., Canup, R. M. & Levison, H. F. On the character and consequences of large impacts in the late stage of terrestrial planet formation. *Icarus* **142**, 219–237 (1999).
4. Greenberg, R., Hartmann, W. K., Chapman, C. R. & Wacker, J. F. Planetesimals to planets—Numerical simulations of collisional evolution. *Icarus* **35**, 1–26 (1978).
5. Weidenschilling, S. J., Spaute, D., Davis, D. R., Marzari, F. & Ohtsuki, K. Accretion evolution of a planetesimals swarm. *Icarus* **128**, 429–455 (1997).
6. Kokubo, E. & Ida, S. Formation of protoplanetary systems and diversity of planetary systems. *Astrophys. J.* **581**, 666–680 (2002).
7. Pierazzo, E. & Melosh, H. J. Hydrocode modelling of oblique impacts: The fate of the projectile. *Meteorit. Planet. Sci.* **35**, 117–130 (2000).
8. Stevenson, D. J. Origin of the moon—The collision hypothesis. *Annu. Rev. Earth Planet. Sci.* **15**, 271–315 (1987).
9. Canup, R. M. Dynamics of lunar formation. *Annu. Rev. Astron. Astrophys.* **42**, 441–475 (2004).
10. Canup, R. & Asphaug, E. Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* **412**, 708–712 (2001).
11. Asphaug, E. & Benz, W. Size, density, and structure of comet Shoemaker-Levy 9 inferred from the physics of tidal breakup. *Icarus* **121**, 225–248 (1996).
12. Jeffreys, H. The relation of cohesion to Roche's limit. *Mon. Not. R. Astron. Soc.* **107**, 260–262 (1947).
13. McKinnon, W. B. & Schenk, P. M. Estimates of comet fragment masses from impact crater chains on Callisto and Ganymede. *Geophys. Res. Lett.* **22**, 1829–1832 (1995).
14. Boss, A. P., Cameron, A. G. W. & Benz, W. Tidal disruption of inviscid planetesimals. *Icarus* **92**, 165–178 (1991).
15. Rettig, T. W., Mumma, M. J., Sobczak, G. J., Hahn, J. M. & DiSanti, M. The nature of Comet Shoemaker-Levy/9 subnuclei from analysis of preimpact Hubble Space Telescope images. *J. Geophys. Res.* **101**, 9271–9281 (1996).
16. Abe, Y., Ohtani, E., Okuchi, T., Righter, K. & Drake, M. in *Origin of the Earth and Moon* (eds Canup, R. & Righter, K.) 413–433 (Univ. Arizona Press, Tucson, 2000).
17. Wilson, L. Relationships between pressure, volatile content and ejecta velocity in three types of volcanic explosion. *J. Volcanol. Geotherm. Res.* **8**, 297–313 (1980).
18. Jeans, J. H. *Problems of Cosmogony and Stellar Dynamics* (Cambridge Univ. Press, Cambridge, 1919).
19. Benz, W. in *The Numerical Modeling of Nonlinear Stellar Pulsations: Problems and Prospects* (ed. Buchler, J. R.) 269–288 (Kluwer Academic, Boston, 1990).
20. Asphaug, E., Agnor, C. & Williams, Q. Tidal forces as drivers of collisional evolution. *Lunar Planet. Sci. Conf. XXXVI* abstr. 2393 (2005); <http://www.lpi.usra.edu/meetings/lpsc2005/pdf/2393.pdf> (2005).
21. Mizuno, H. & Boss, A. P. Tidal disruption of dissipative planetesimals. *Icarus* **63**, 109–133 (1985).
22. Sridhar, S. & Tremaine, S. Tidal disruption of viscous bodies. *Icarus* **95**, 86–99 (1992).
23. Grady, D. E. & Kipp, M. E. Continuum modeling of explosive fracture in oil shale. *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **17**, 147–157 (1980).
24. Asphaug, E., Ryan, E. & Zuber, M. in *Asteroids III* Table I (eds Bottke, W. F. Jr, Cellino, A., Paolicchi, P. & Binzel, R. P.) 463–484 (Univ. Arizona Press, Tucson, 2002).
25. Burbine, T. H., Meibom, A. & Binzel, R. P. Mantle material in the main belt: Battered to bits? *Meteoritics* **31**, 607–620 (1996).
26. Mei, S., Bai, W., Hiraga, T. & Kohlstedt, D. L. Influence of melt on the creep behaviour of olivine-basalt aggregates under hydrous conditions. *Earth Planet. Sci. Lett.* **201**, 491–507 (2002).
27. Walzer, U., Hendel, R. & Baumgardner, J. The effects of a variation of the radial viscosity profile on mantle evolution. *Tectonophysics* **384**, 55–90 (2004).
28. Tackley, P. J. Convection in Io's asthenosphere: Redistribution of nonuniform tidal heating by mean flows. *J. Geophys. Res.* **106**, 32971–32982 (2001).
29. Wetherill, G. W. An alternative model for the formation of the asteroids. *Icarus* **100**, 307–325 (1992).
30. Bottke, W. F. *et al.* The fossilized size distribution of the main asteroid belt. *Icarus* **175**, 111–140 (2005).
31. McCoy, T. J. *et al.* A petrologic and isotopic study of lodranites: Evidence for early formation as partial melt residues from heterogeneous precursors. *Geochim. Cosmochim. Acta* **61**, 623–637 (1997).

32. Haack, H., Scott, E. R. D. & Rasmussen, K. L. Thermal and shock history of mesosiderites and their large parent asteroid. *Geochim. Cosmochim. Acta* **60**, 2609–2619 (1996).
33. Wilson, L., Keil, K., Browning, L. B., Krot, A. N. & Bourcier, W. Early aqueous alteration, explosive disruption, and re-processing of asteroids. *Meteorit. Planet. Sci.* **34**, 541–557 (1999).
34. Keil, K., Stöffler, D., Love, S. G. & Scott, E. R. D. Constraints on the role of impact heating and melting in asteroids. *Meteoritics* **32**, 349–363 (1997).
35. Ostro, S. J. *et al.* Radar observations of asteroid 216 Kleopatra. *Science* **288**, 836–839 (2000).
36. Rivkin, A. S., Howell, E. S., Lebofsky, L. A., Clark, B. E. & Britt, D. T. The nature of M-class asteroids from 3- μ m observations. *Icarus* **145**, 351–368 (2000).
37. Davis, D. R., Chapman, C. R., Greenberg, R. & Weidenschilling, S. J. Collisional history of asteroids: Evidence from Vesta and the Hirayama families. *Icarus* **62**, 30–53 (1985).
38. Dohnanyi, J. W. Collisional models of asteroids and their debris. *J. Geophys. Res.* **74**, 2531–2554 (1969).
39. Binzel, R. P. *et al.* Geologic mapping of Vesta from 1994 Hubble Space Telescope images. *Icarus* **128**, 95–103 (1997).
40. Greenwood, R. C., Franchi, I. A., Jambon, A. & Buchanan, P. C. Widespread magma oceans on asteroidal bodies in the early Solar System. *Nature* **435**, 916–919 (2005).
41. Rivers, M. L. & Carmichael, I. S. E. Ultrasonic studies of silicate melts. *J. Geophys. Res.* **92**, 9247–9270 (1987).
42. Ochs, F. A. III & Lange, R. A. The density of hydrous magmatic liquids. *Science* **283**, 1314–1317 (1999).
43. Dixon, J. E., Stolper, E. M. & Holloway, J. R. An experimental study of water and carbon dioxide solubilities in mid-ocean ridge basaltic liquids. Part I. Calibration and solubility models. *J. Petrol.* **36**, 1607–1631 (1995).
44. Benz, W. & Asphaug, E. Catastrophic disruptions revisited. *Icarus* **142**, 5–20 (1999).

Acknowledgements This research was sponsored by NASA's Planetary Geology and Geophysics Program, "Small Bodies and Planetary Collisions". We benefited from discussions with a number of colleagues, including W. F. Bottke and R. Canup. We particularly thank D. Stevenson and K. Zahnle for comments on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.A. (asphaug@pmc.ucsc.edu).

Widespread amphibian extinctions from epidemic disease driven by global warming

J. Alan Pounds¹, Martín R. Bustamante², Luis A. Coloma², Jamie A. Consuegra³, Michael P. L. Fogden¹, Pru N. Foster^{4†}, Enrique La Marca⁵, Karen L. Masters⁶, Andrés Merino-Viteri², Robert Puschendorf⁷, Santiago R. Ron^{2,8}, G. Arturo Sánchez-Azofeifa⁹, Christopher J. Still¹⁰ & Bruce E. Young¹¹

As the Earth warms, many species are likely to disappear, often because of changing disease dynamics. Here we show that a recent mass extinction associated with pathogen outbreaks is tied to global warming. Seventeen years ago, in the mountains of Costa Rica, the Monteverde harlequin frog (*Atelopus* sp.) vanished along with the golden toad (*Bufo periglenes*). An estimated 67% of the 110 or so species of *Atelopus*, which are endemic to the American tropics, have met the same fate, and a pathogenic chytrid fungus (*Batrachochytrium dendrobatidis*) is implicated. Analysing the timing of losses in relation to changes in sea surface and air temperatures, we conclude with 'very high confidence' (>99%, following the Intergovernmental Panel on Climate Change, IPCC) that large-scale warming is a key factor in the disappearances. We propose that temperatures at many highland localities are shifting towards the growth optimum of *Batrachochytrium*, thus encouraging outbreaks. With climate change promoting infectious disease and eroding biodiversity, the urgency of reducing greenhouse-gas concentrations is now undeniable.

Humans are altering the Earth's climate^{1–4} and thus the workings of living systems^{5–8}, including pathogens and their hosts^{9–11}. Among the predicted outcomes is the extinction of many species^{10,12}, but detecting such an effect is difficult against a backdrop of other changes, especially habitat destruction. One approach is to focus on organisms for which current rates of extinction exceed those expected from habitat loss. Amphibians are a case in point. Thousands of species have declined, and hundreds are on the brink of extinction or have already vanished¹³. The Global Amphibian Assessment (GAA) lists 427 species as "critically endangered", including 122 species that are "possibly extinct"¹³. A majority of the former, and nearly all of the latter, have declined even in seemingly undisturbed environments.

The causes have remained unclear, in part because of their complexity^{14–16}. Although pathogens are implicated^{14–28}, their relationship to environmental change is poorly understood. Here we test the "climate-linked epidemic hypothesis"^{29–34}, which predicts declines in unusually warm years but does not assume a particular disease or chain of events. Recent studies have considered this idea^{15,18,21,23,28}, yet data have not permitted a geographically broad test that examines landscape alteration, global warming and climate fluctuations on the timescale of El Niño. Suffering widespread extinctions often despite habitat protection, harlequin frogs (*Atelopus*) afford such a test. A new database, produced by 75 researchers, documents the case in unprecedented detail, owing to the nature of

these members of the toad family (Bufonidae)²⁶. Brightly coloured and active during the day near streams, most are readily observed and identified. For the first time, data indicate when each of numerous species was seen for the last time.

Our analyses capitalise on insights gained by alternating between large and small spatial scales³⁵ (Supplementary Fig. 1). Since epidemics of *Batrachochytrium* are implicated in *Atelopus* extinctions in Central and South America²⁶, we first explain that the predicted association with warm years, if juxtaposed with theory regarding this chytrid, is a paradox. We then: (1) assess large-scale altitudinal patterns of extinction risk with this paradox in mind; (2) consider determinants of local climate in the case of the golden toad and the Monteverde harlequin frog to select large-scale temperature signals for analysing the biological patterns; (3) show that the timing of the widespread extinctions is strongly tied to these signals; and (4) explore local climate from a chytrid's viewpoint to frame a solution to the paradox.

The climate-chytrid paradox

The climate-linked epidemic hypothesis predicts amphibian declines in unusually warm years, because shifts in temperature or related variables often influence disease dynamics^{9–11}. As temperatures rise, climate fluctuations may cross thresholds for certain pathogens, triggering outbreaks. Many diseases are expected to become more lethal, or to spread more readily, as the Earth warms^{9–11}.

¹Golden Toad Laboratory for Conservation, Monteverde Cloud Forest Preserve and Tropical Science Center, Santa Elena, Puntarenas 5655-73, Costa Rica. ²Museo de Zoología, Centro de Biodiversidad y Ambiente, Escuela de Biología, Pontificia Universidad Católica del Ecuador, Avenida 12 de Octubre 1076 y Roca, Apartado 17-01-2184, Quito, Ecuador. ³Department of Environmental Science, Barnard College, Columbia University, 3009 Broadway, New York, New York 10027, USA. ⁴Center for Climate Studies Research, University of Tokyo, Komaba, 4-6-1, Meguro-ku, Tokyo 153-8904, Japan. ⁵Laboratorio de Biogeografía, Escuela de Geografía, Facultad de Ciencias Forestales y Ambientales, Universidad de Los Andes, Apartado 116, Mérida 5101-A, Venezuela. ⁶Council for International Educational Exchange, Monteverde, Puntarenas 5655-26, Costa Rica. ⁷Escuela de Biología, Universidad de Costa Rica, San Pedro, Costa Rica. ⁸Texas Memorial Museum and Department of Integrative Biology, University of Texas, Austin, Texas 78712, USA. ⁹Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta T6G 2E3, Canada. ¹⁰Department of Geography, 3611 Ellison Hall, University of California at Santa Barbara, Santa Barbara, California 93106, USA. ¹¹NatureServe, Monteverde, Puntarenas 5655-75, Costa Rica. †Present address: Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK.

Chytridiomycosis, caused by *Batrachochytrium*, is thought to be an exception¹⁰. This chytrid grows on amphibian skin and produces aquatic zoospores^{22,24}. Widespread and ranging from deserts and lowland rainforests to cold mountain tops²⁷, it is sometimes a non-lethal parasite and possibly a saprophyte^{19,25}. It is associated with host mortality in highlands or during winter²², and, according to theory, becomes more pathogenic at lower temperatures^{19,22}. Hence, the idea that it causes declines in warm years is paradoxical. Moreover, the fungus is apparently more lethal under moist conditions^{24,26}, yet, at many affected sites, warm years are comparatively dry.

Ideas of two sorts could resolve this paradox. First, warm or dry conditions may stress amphibians, possibly increasing susceptibility to disease²⁹. Second, warm years could favour *Batrachochytrium* directly. The prevailing idea—that lower temperatures benefit the chytrid^{19,22}—might be an oversimplification of the pathogen's response to climate.

Altitudinal patterns of extinction risk

This prevailing idea predicts greater extinction risk for higher-elevation species. Many are already prone to extinction, because geographic ranges tend to decrease in size with increasing elevation. The probability of disappearance might thus be expected to increase from lowlands to mountain tops.

For a preliminary test with *Atelopus*, we consider 100 species for which data indicate the last year of observation (LYO). We recognize two tiers. According to La Marca *et al.* (ref. 26), the population data are sufficient to judge whether tier-one species ($n = 51$) have declined, but not tier-two species ($n = 49$). Throughout our analyses, patterns are similar for tier one and for tiers one and two combined. Adding tier two increases error but provides insights. We score species as having disappeared if the LYO is 1998 or earlier.

The altitudinal patterns are more complex than expected (Fig. 1). Using a sliding window to assess how the probability of disappearance varies with species' lower elevational limit, we find three breakpoints. The percentage of species lost increases sharply at 200 m, and again at 1,000 m. It decreases, however, at 2,400 m, and thus peaks at middle elevations, suggesting that low temperatures as

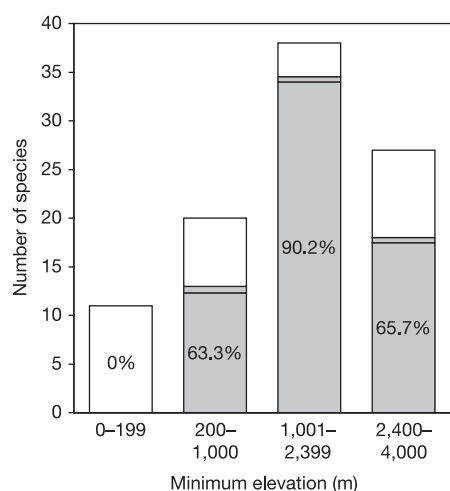


Figure 1 | Altitudinal patterns in the *Atelopus* extinctions. Bars indicate the number of species known per altitudinal zone (total $n = 96$), and the grey-shaded portions represent the estimated percentage of species lost from each. This percentage differs among zones ($\chi^2 = 31.4$, degrees of freedom = 3, $P < 0.0001$), but not between the tier-one species-set and the species-set that combines tiers one and two (Fisher exact tests, $P > 0.9$). The double lines indicate the values for each of these two species-sets; the percentage labels are the averages of the two. The percentage for the zone affected most severely differs from that of each adjacent zone (Fisher exact tests, $P < 0.036$).

well as high ones may limit the impact of *Batrachochytrium*. The altitudinal effects remain significant when we control for range size, which also influences extinction probability. Average range size decreases from lower to higher zones as defined in Fig. 1, but is similar for the upper two.

These altitudinal patterns contribute to the severity of losses. For instance, the zone losing the highest percentage of species had the greatest diversity (Fig. 1). Our overall estimate that 67% of the species have disappeared is weighted by the number of species per zone. Although extinction probabilities are independent of tier, an unweighted estimate based on tier one alone (57%) under-represents the severely affected mid-elevation species.

GAA data for New World amphibians¹³ suggest similar altitudinal patterns (Supplementary Fig. 2). The percentage of species extinct or threatened is largest at middle elevations, even though higher-elevation species generally have smaller ranges. Clearly, the role of climate needs re-evaluating.

Temperature signals

To select temperature signals, we consider the scale at which local climates are determined. In Costa Rica's Monteverde cloud forest, reduced mist frequency in warm years is associated with shifts in populations of birds, reptiles and amphibians, including the disappearance of the golden toad and the Monteverde harlequin frog³¹. Whereas nearby lowland deforestation might have influenced conditions³⁶, temperatures in Central and South America agree with simulated responses to greenhouse-gas accumulation³. Here we quantify the extent and timing of deforestation upwind of Monteverde, model regional climate, and consider how local trends relate to sea surface temperature (SST) and air temperature (AT) on varying scales.

We focus on warming and the growing number of dry days, which reflects increasing precipitation variability and declining mist frequency³¹. While the latter probably affects many organisms, impacts of correlated climatic changes are hard to separate, and not all *Atelopus* extinctions have occurred in habitats where mist is vital²⁶. In any case, large-scale temperature shifts, integrating various aspects of climate change, are a likely common denominator.

The chiefly historical deforestation probably enhanced sensitivity to warming but cannot easily explain the trends. Using LANDSAT images and aerial photos, we assess changes in a 35-km-wide belt representing the trade-wind path from the Caribbean shore to the 500-m contour. Clearing through the year 2000 claimed about 38% of this belt. The loss, however, was only 11% during 1975–2000, when the changes occurred at Monteverde, and 9% during 1960–1975. The area of the San Carlos Plain directly upwind was cleared before 1940 (ref. 36).

In contrast, global temperatures have climbed steeply since the early 1970s (refs 1–4). In the tropics, all forest regions have warmed^{37,38}, and mountain glaciers are rapidly melting³⁹. During 1975–2000, SST and AT for the tropics, both of which are averages for 30° N–30° S, were highly correlated (Fig. 2a). The latter increased by 0.18 °C per decade, which is triple the average rate of warming for the twentieth century. It is 18 times the inferred average rate for a mid-elevation cloud forest in the Andes during the 8,000-year transition from the ice ages to modern times (Pleistocene–Holocene)⁴⁰. It is similarly more rapid than the non-directional changes of the preceding 30,000 years.

The recent warming, our work suggests, has reduced mist frequency at Monteverde by raising heights of orographic cloud formation. These altitudes depend on relative humidity in the trade winds ascending the mountain slopes, and thus on moisture content and temperature⁴¹. In our simulations, large-scale warming reduces relative humidity locally much more than the observed deforestation (Supplementary Fig. 3). The growing number of dry days is consistent. It is correlated with SST in each of six regions: offshore Caribbean (near Costa Rica), offshore Pacific, equatorial

Pacific (Niño-3 region), deep tropics (10°N–10°S), tropics, and the globe. A residual trend remains, however, unless we consider the tropics or the globe (Supplementary Fig. 4). Large-scale and local climatic changes are strikingly concordant, with fluctuations related to El Niño superimposed on the trends (Fig. 2b, c). Thus, analysis of AT or SST for the tropics should capture temperature shifts influencing the relevant ecological processes.

Signatures of warming

Accordingly, the biological changes at Monteverde are associated statistically with AT and SST for the tropics but not with Niño-region SST alone (Supplementary Fig. 5). Correlations are evident for the shift of lower-elevation breeding birds up the mountain slopes, and for the decline of highland lizards. Likewise, the episodic losses of amphibians occurred in years that were unusually warm across the tropics. To examine the overall biological pattern in relation to AT for the tropics, we randomize this signal for 1979–1998. In each of 10,000 iterations, we assign the annual means at random to the 20 years and recalculate indices of association. The results confirm that none of the observed relationships is likely to have arisen by chance (Supplementary Table 1). Moreover, in only one iteration are the modelled values all as extreme as the observed ones, indicating a high probability that large-scale warming is affecting local ecology.

We examine the timing of the widespread *Atelopus* extinctions in relation to the same temperature signals. The Jambato toad (*Atelopus ignescens*) of Ecuador and the Monteverde harlequin frog suggest the working hypothesis that species tend to be seen for the last time right after a relatively warm year^{21,29,31}. Both were last found in 1988, following a temperature peak in 1987. Before 1988, the Jambato toad was present during 64% of visits to sites throughout its 10,234-km² range²¹. After 1988, it was absent at all sites, implying synchronous declines across localities. The degree of synchrony, however, differs

among species, and survivors of population crashes persist for variable lengths of time. Furthermore, the spatial and temporal coverage of sampling varies, introducing error.

At any rate, the climate-linked epidemic hypothesis predicts an association between disappearances and warm years, but not a one-to-one correspondence³³. An *Atelopus* population might survive despite warm weather if the pathogens are absent from particular sites within their range, or if they are spreading but have not reached certain areas. Factors that discourage pathogen transmission, such as low host-density, may likewise forestall declines¹⁰. Although temperature shifts can entrain multiple outbreaks, fast-moving waves of infection might also synchronize declines across localities, and an *Atelopus* population experiencing normal weather could succumb to a wave set in motion elsewhere.

Despite the potential variability, the extinctions show signatures of warming. Like the biological changes at Monteverde, they are associated statistically with AT and SST for the tropics but not with Niño-region SST alone (Fig. 3 & Supplementary Table 2). Around 80% of the species that have disappeared were seen for the last time right after a relatively warm year. For tier one, and for tiers one and two combined, we use Monte Carlo methods to generate 10,000 random frequency distributions for comparison with the observed distributions. These analyses confer 'very high confidence' (>99%, following the IPCC^{1,5}) that the tendencies are not due to chance and that large-scale warming is a key factor. Patterns for the two species-sets are comparable despite differences in error rate and time period. Spanning a longer period, the combined data indicate depletion of the most vulnerable species by the late 1990s.

Results are consistent when we repeat the analyses with various subsets of species to consider sources of variation, error and uncertainty (Supplementary Table 2). For instance, because some undescribed species are poorly known, we repeat the analyses including only described ones. Likewise, we consider occurrence in protected

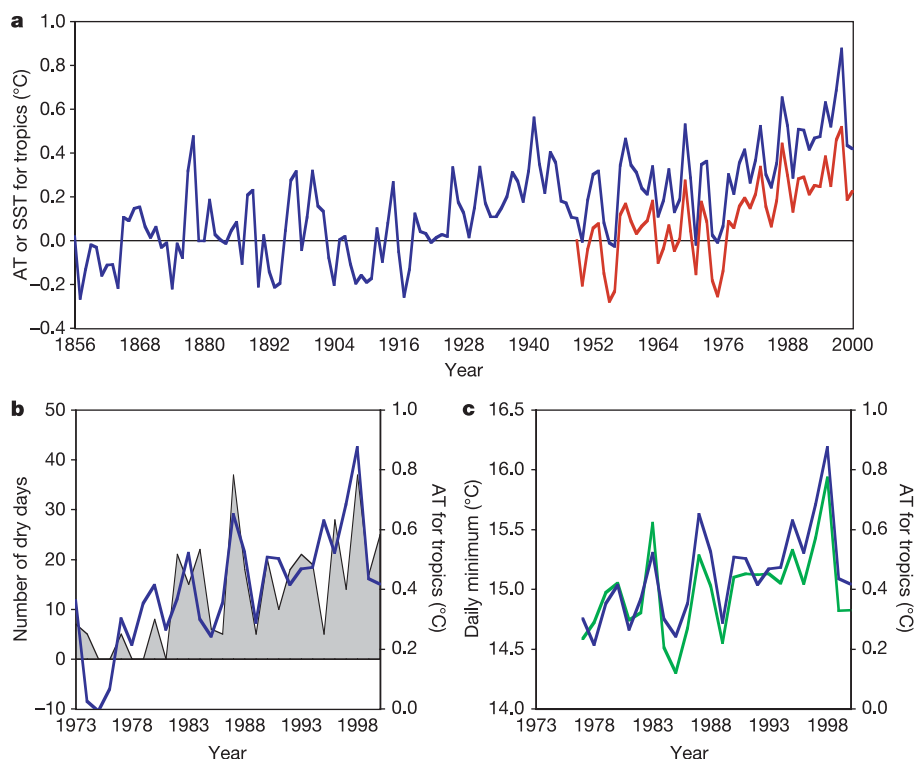


Figure 2 | AT and SST for the tropics and their relationship to climatic trends at Monteverde. AT for the tropics (blue line) is correlated with: **a**, SST for the tropics (red line) ($r = 0.97$, $P < 0.0001$, $n = 51$); **b**, number of dry days in runs ≥ 5 days (grey-shaded area) ($r = 0.70$, $P < 0.0001$, $n = 28$);

and **c**, local daily minimum AT (green line) ($r = 0.91$, $P < 0.0001$, $n = 24$). Temperatures are annual averages. AT and SST for the tropics are departures from a baseline mean (for 1856–1895 and 1951–1979, respectively).

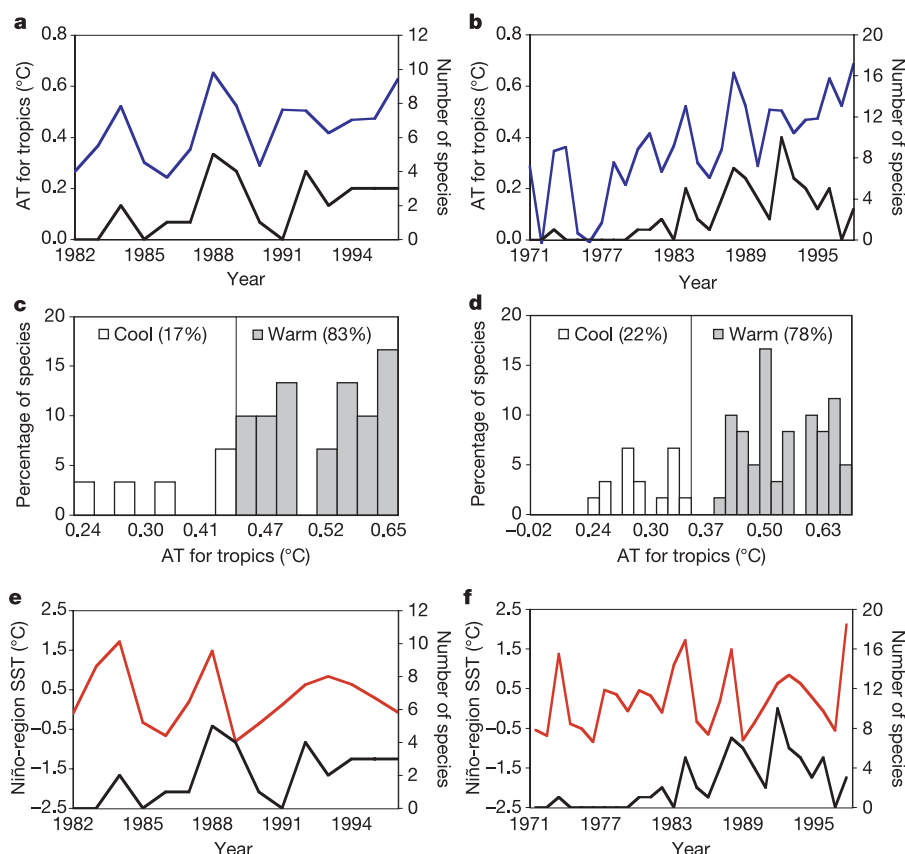


Figure 3 | Signatures of warming in the *Atelopus* extinctions. For tier one (a) and tiers one and two combined (b), the number of species observed for the last time (black line) is related to AT for the tropics in the preceding year (blue line). c, d, Percentage of species observed for the last time following a relatively warm year exceeds that expected by chance (c, tier one, 83%,

$P < 0.002$, $n = 29$; d, tiers one and two, 78%, $P < 0.0001$, $n = 68$). e, f, The same patterns (black lines as in a, b) are not significantly related to Niño-region SST (red line). The corresponding percentages do not exceed those expected by chance (e, tier one, 55%, $P > 0.43$; f, tiers one and two, 62%, $P > 0.12$). Temperatures are calculated as in Fig. 2.

areas, accessibility of regions, and factors that might influence the timing of extinction. Several analyses provide independent tests of our hypothesis. The strength of association between warm years and disappearances is not related to altitude, latitude or range size. Accordingly, conclusions are similar for 'northern' and 'southern' species, and for 'higher-elevation' and 'lower-elevation' ones.

In these analyses, AT or SST for the tropics serves as a relative index, often registering smaller shifts than local indices. In 1987, the former averaged 0.65°C above the baseline (Fig. 2), whereas local temperatures relevant to the Jambato toad's extinction in the highlands of Ecuador were almost 2.0°C above a century-long mean²¹. The difference may reflect, in part, increasing atmospheric moisture, which can amplify the signal at higher altitudes^{42,43}. Global warming accelerates evaporation and raises the air's capacity to hold water. As water vapour rises and condenses, latent heat is transferred to the atmosphere.

A climate for chytrids

Increased water vapour can also translate into enhanced cloud cover—often with the help of condensation nuclei from particulate air pollution (aerosols)⁴⁴—creating additional feedbacks that influence surface temperatures⁴⁵. This may be notable in places where air rises strongly, such as in mountainous regions. Reducing heat loss at night, cloud cover adds to nocturnal warming. By impeding solar radiation, however, it moderates daytime trends and may reverse them. In many areas, the daily temperature range is declining as the minimum rises faster than the maximum⁴⁵.

Such trends are evident in the highlands of Central and South

America^{46–48}. At Monteverde, regardless of the season, the daily minimum is rising while the daily maximum is falling (Fig. 4a). For 11 Colombian and Venezuelan stations with quality-checked, long-term data⁴⁷, we compare 1941–1970, preceding the *Atelopus* extinctions, to the first decade with major losses (1981–1990). The minimum again shows an increase and the maximum a decrease (Fig. 4b). At some localities both are rising, but the former disproportionately so^{46–48}. These trends imply increasing cloud cover that contributes to warming at night but diminishes it during the day⁴⁵.

Cloudiness should favour the chytrids. These fungi reportedly grow best at 17 to 25°C , peaking at 23°C (refs 19, 22, 24). They stop growing at 28°C and die at 30°C . Shielding them from excessive warmth and fostering moist conditions, cloud cover may promote their survival, growth and reproduction. At Monteverde, where ambient daytime temperatures are usually chytrid-friendly, temperatures inside sunlit moss mats, bromeliads or leaf litter often exceed 30°C (ref. 34). Cloudiness, however, largely shuts down radiant heating, forcing thermal environments to mirror ambient conditions. Microscale trends can thus dwarf local ambient trends, which, in some places, might even be of opposite sign. If amphibians seek warmth to combat infection, increasing cloudiness might hamper their defences³⁴. In any case, local or microscale cooling should often benefit the chytrids.

So why should these pathogens flourish in the highlands during warm years? The answer, we suggest, lies in the difference between night and day. To consider this difference, we plot daily minimum and maximum temperatures in relation to altitude for 50 localities, from Costa Rica to Peru, along with the optimal temperatures for

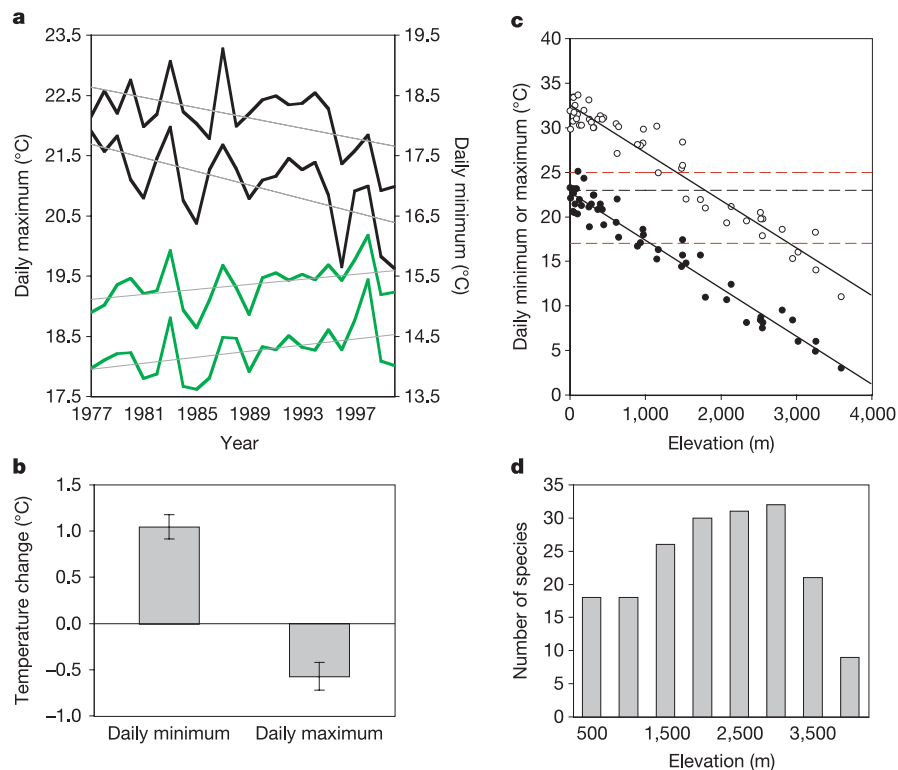


Figure 4 | Daily minimum and maximum temperatures and the chytrid-thermal-optimum hypothesis. **a**, At Monteverde, the average daily minimum (green lines) and maximum (black lines) for warmer months (March–October) and cooler months (November–February) show trends ($|r| \geq 0.43$, $P < 0.038$, $n = 24$). **b**, In Colombia and Venezuela, station averages during the extinctions differed from those of earlier decades. Values

are mean changes ($\pm$ s.e.m.). **c**, For Costa Rica to Peru, annual average daily minimum (closed circles) and maximum (open circles) vary by altitude. Dashed lines give reported optimal temperature range (red dashed line) and optimum temperature (black dashed line) for *Batrachochytrium*. **d**, Number of *Atelopus* species per 500-m belt is labelled by upper limit. Six ranged above 4,000 m.

growth of *Batrachochytrium* (Fig. 4c). We also plot the altitudinal distribution of *Atelopus* (Fig. 4d). Two patterns are clear. First, just as the lowlands are often too warm for the chytrids during the day, the highlands are often too cool for them at night. Second, most *Atelopus* extinctions have occurred at elevations where the minimum temperature is shifting towards the growth optimum for these pathogens. Thus, we propose the chytrid-thermal-optimum hypothesis, in which daytime cooling (local or microscale) and night time warming accelerate disease development. The impacts at night may explain the association with warm years and thereby resolve the climate–chytrid paradox.

Conclusions

We establish that global climate change is already causing the extinction of species. Taking our results and recent findings that tie the same losses to disease, we conclude that climate-driven epidemics are an immediate threat to biodiversity. Our study sheds light on the amphibian–decline mystery by showing that large-scale warming is a key factor. It also points to a chain of events whereby this warming may accelerate disease development by translating into local or microscale temperature shifts—increases and decreases—favourable to *Batrachochytrium*. The case illustrates how greenhouse warming and the resultant intensification of the hydrological cycle, together with aerosol pollution, may affect life on Earth. Influencing patterns of cloud formation, these agents alter the thermal, light and moisture environments of many organisms, changing ecological interactions and threatening species survival.

METHODS

Assessing extinction probability in relation to altitude. We examine the influence of altitudinal distribution while considering range size. Data indicate

the last year of observation (LYO) for 104 of the 113 species of *Atelopus* (see Appendix A in the Supplementary Information). Focusing on recent losses, we exclude four species known only from historical records (1950 or earlier). Variables include minimum and maximum elevations²⁶, elevational midpoint and longest axis of the range polygon (GAA data¹³). For statistical analyses, we use logistic regression and contingency tables. Sample sizes vary because some data are missing, particularly for undescribed species. Although extinction probability is related to each altitudinal variable, we focus on minimum elevation, because it indicates whether populations occur at low elevations, which may provide a refuge from chytridiomycosis. Ranking species by minimum elevation, from lowest to highest, we use a sliding window to compare extinction probabilities between successive subsets of ten, shifting the window one species at a time. The first ten encompass the altitudes (<200 m) at which no species has disappeared.

Selecting temperature signals. The regional modelling generates climate-change scenarios, which we test by analysing local trends in relation to temperature on varying spatial scales. In the simulations, we prescribe realistic partial clearing, based on the remote-sensing data. We compare its impact to that of an increase in AT and SST approximating tropical warming over 1973–2000 and, in a separate run, the 1986–1987 El Niño. We examine aerial photos from Costa Rica's National Geographic Institute for 1960, LANDSAT Multispectral Scanner images for 1975, and LANDSAT Thematic Mapper 7 images for the year 2000. Detecting forest fragments $\geq 0.03 \text{ km}^2$, and classifying areas as forest if canopy density is $\geq 80\%$, the techniques⁴⁹ produce land-cover maps at a scale of 1:250,000. We run our simulations with the Regional Atmospheric Modelling System (RAMS)⁵⁰ at a maximum horizontal resolution of 1.6 km. To prescribe initial and boundary conditions, we use the ECMWF Global Reanalysis (2.5°-resolution ERA-15 Pressure Level Analysis) from the European Centre for Medium-range Weather Forecasts. The GDTPO30 digital elevation model from the US Geological Survey defines terrain. To examine local trends in relation to temperature on varying scales, we examine SST data (2°-resolution) from the National Oceanic and Atmospheric Administration's (NOAA's) National Centers for Environmental Prediction (NCEP), and AT data (5°-resolution) from the Climate Research Unit (CRU), University of East

Anglia. The local data, including tallies of dry days for January–May, are from 1,540 m on Monteverde's upper Pacific slope³¹.

Testing for a link to global warming. We use resampling methods to analyse biological patterns in relation to large-scale temperature signals. 'Warm years' are above average for the period of analysis. For Monteverde, we examine 1979–1998, which encompasses the field observations³¹. To produce a single time series for anoline lizards, we average data for the two declining species, which are highly correlated. The *Atelopus* data are from various independent studies²⁶. To prevent bias, persons contributing or compiling these data were not told how they would be analysed in relation to climate. We consider AT for the tropics over the last year of observation (LYO) of each species, for one, two and three years before, and for averages across these years. The period of analysis is defined accordingly. We find an association, however, only for one year before. To examine the strength of this association in relation to altitudinal and latitudinal distribution and range size, we use logistic regression. Latitudinal variables include northern and southern limits and range midpoint. The analyses of biological patterns in relation to Niño-region SST (departures relative to 1950–1979) yield similar conclusions regardless of the signal examined: Niño-1 and -2, Niño-3, these combined, or Niño-1 and -2 combined with Niño-3 and -4. We present results for the latter, composite, signal.

Resolving the climate–chytrid paradox. We focus on temperature, since altitudinal patterns in the declines underscore its importance and because the extinctions are strongly associated with it. Temperature shifts are presumably more coherent spatially than attendant climatic changes. Our premise is that any pathogen with an optimal range for growth and reproduction will be sensitive to low temperatures as well as high ones, as suggested by the altitudinal patterns of extinction risk. (see the Supplementary Notes pertaining to "A climate for chytrids", which explores the meaning of the observed breakpoints.) Comparing temperatures for the warmer and cooler months at Monteverde shows that, seasonally, night time temperatures in the highland tropics often lie even farther below the optimal range for *Batrachochytrium* than average conditions suggest. To examine daily minimum and maximum temperatures for the 11 Colombian and Venezuelan stations, we use the published averages for particular decades⁴⁷. The 50 localities from Costa Rica to Peru mostly represent inland areas. The corresponding analyses do not control for latitude or sampling period, yet provide a generalized altitudinal profile of minimum and maximum temperatures. Numbers of *Atelopus* species that inhabited the different altitudinal zones are based on $n = 96$. Some inhabited more than one zone.

Received 2 June; accepted 21 September 2005.

- Houghton, J. T. et al. (eds) *Climate Change 2001, The Scientific Basis. Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge, 2001).
- Santer, B. D. et al. Contributions of anthropogenic and natural forcing to recent tropopause height changes. *Science* **301**, 479–483 (2003).
- Stott, P. A. Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.* **30**, 1728–1731 (2003).
- Barnett, T. P. et al. Penetration of human-induced warming into the world's oceans. *Science* **309**, 284–287 (2005).
- Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
- Root, T. L. et al. Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57–60 (2003).
- Root, T. L., MacMynowski, D. P., Mastrandrea, M. D. & Schneider, S. H. Human-modified temperatures induce species changes: Joint attribution. *Proc. Natl Acad. Sci. USA* **102**, 7465–7469 (2005).
- Lovejoy, T. & Hannah, L. (eds) *Climate Change and Biodiversity* (Yale Univ. Press, New Haven, Connecticut, 2005).
- Epstein, P. R. Climate change and emerging infectious diseases. *Microbes Infect.* **3**, 747–754 (2001).
- Harvell, C. D. et al. Climate warming and disease risks for terrestrial and marine biota. *Science* **296**, 2158–2162 (2002).
- Rodó, X., Pascual, M., Fuchs, G. & Faruque, A. S. G. ENSO and cholera: A nonstationary link related to climate change? *Proc. Natl Acad. Sci. USA* **99**, 12901–12906 (2002).
- Thomas, C. D. et al. Extinction risk from climate change. *Nature* **427**, 145–148 (2004).
- Stuart, S. M. et al. Status and trends of amphibian declines and extinctions worldwide. *Science* **306**, 1783–1786 (2004).
- Alford, R. A. & Richards, S. J. Global amphibian declines: a problem in applied ecology. *Annu. Rev. Ecol. Syst.* **30**, 133–165 (1999).
- Kiesecker, J. M., Blaustein, A. R. & Belden, L. K. Complex causes of amphibian population declines. *Nature* **410**, 681–684 (2001).
- Blaustein, A. R. & Kiesecker, J. M. Complexity in conservation: lessons from the global decline of amphibian populations. *Ecol. Lett.* **5**, 597–608 (2002).
- Berger, L. et al. Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proc. Natl Acad. Sci. USA* **95**, 9031–9036 (1998).
- Carey, C. & Alexander, M. A. Climate change and amphibian declines: Is there a link? *Divers. Distrib.* **9**, 111–121 (2003).
- Daszak, P., Cunningham, A. A. & Hyatt, A. D. Infectious disease and amphibian population declines. *Divers. Distrib.* **9**, 141–150 (2003).
- Lips, K. R., Green, D. E. & Pappendick, R. Chytridiomycosis in wild frogs from southern Costa Rica. *J. Herpetol.* **37**, 215–218 (2003).
- Ron, S. R., Duellman, W. E., Coloma, L. A. & Bustamante, M. Population decline of the Jambato toad *Atelopus ignescens* (Anura: Bufonidae) in the Andes of Ecuador. *J. Herpetol.* **37**, 116–126 (2003).
- Berger, L. et al. Effect of season and temperature on mortality in amphibians due to chytridiomycosis. *Aust. Vet. J.* **82**, 434–439 (2004).
- Burrowes, P. A., Joglar, R. L. & Green, D. E. Potential causes for amphibian declines in Puerto Rico. *Herpetologica* **60**, 141–154 (2004).
- Piotrowski, J. S., Annis, S. L. & Longcore, J. E. Physiology of *Batrachochytrium dendrobatidis*, a chytrid pathogen of amphibians. *Mycologia* **96**, 9–15 (2004).
- Retallack, R. W. R., McCallum, H. & Speare, R. Endemic infection of the amphibian chytrid fungus in a frog community post-decline. *PLoS Biol.* **2**, 1966–1971 (2004).
- La Marca, E. et al. Catastrophic population declines and extinctions in Neotropical harlequin frogs (Bufonidae: *Atelopus*). *Biotropica* **37**, 190–201 (2005).
- Ron, S. R. Predicting the distribution of the amphibian pathogen *Batrachochytrium dendrobatidis* in the New World. *Biotropica* **37**, 209–221 (2005).
- Merino-Viteri, A., Coloma, L. A. & Almandáriz, A. in *Studies on the Andean Frogs of the Genera Telmatobius and Brachophryne* (eds Lavilla, E. O. & De la Riva, I.) 9–37 (Asociación Herpetológica Española, Valencia, 2005).
- Pounds, J. A. & Crump, M. L. Amphibian declines and climate disturbance: the case of the golden toad and the harlequin frog. *Conserv. Biol.* **8**, 72–85 (1994).
- Pounds, J. A., Fogden, M. P. L., Savage, J. M. & Gorman, G. C. Tests of null models for amphibian declines on a tropical mountain. *Conserv. Biol.* **11**, 1307–1322 (1997).
- Pounds, J. A., Fogden, M. P. L. & Campbell, J. H. Biological response to climate change on a tropical mountain. *Nature* **398**, 611–615 (1999).
- Pounds, J. A. in *Monteverde: Ecology and Conservation of a Tropical Cloud Forest* (eds Nadkarni, N. M. & Wheelwright, N. T.) 149–177 (Oxford Univ. Press, Oxford, 2000).
- Pounds, J. A. Climate and amphibian declines. *Nature* **410**, 639–640 (2001).
- Pounds, J. A. & Puschendorf, R. Ecology: Clouded futures. *Nature* **427**, 107–109 (2004).
- Root, T. L. & Schneider, S. H. Ecology and climate: Research strategies and implications. *Science* **269**, 334–341 (1995).
- Lawton, R. O., Nair, U. S., Pielke, R. A. Sr & Welch, R. M. Climatic impact of tropical lowland deforestation on nearby montane cloud forests. *Science* **294**, 584–587 (2001).
- Vuille, M., Bradley, R. S., Werner, M. & Keimig, F. 20th century climate change in the tropical Andes: observations and model results. *Clim. Change* **59**, 75–99 (2003).
- Malhi, Y. & Wright, J. Spatial patterns and recent trends in the climate of tropical forest regions. *Proc. R. Soc. Lond. B* **359**, 311–329 (2004).
- Thompson, L. G. et al. Tropical glacier and ice core evidence of climate change on annual to millennial time scales. *Clim. Change* **59**, 137–155 (2003).
- Bush, M. B., Silman, M. R. & Urrego, D. H. 48,000 years of climate and forest change in a biodiversity hot spot. *Science* **303**, 827–829 (2004).
- Still, C. J., Foster, P. N. & Schneider, S. H. Simulating the effects of climate change on tropical montane cloud forests. *Nature* **398**, 608–610 (1999).
- Wentz, F. J. & Schabel, M. Precise climate monitoring using complementary satellite data sets. *Nature* **403**, 414–416 (2000).
- Ross, R. J. & Elliott, W. P. Radiosonde-based Northern Hemisphere tropospheric water vapor trends. *J. Clim.* **14**, 1602–1612 (2001).
- Penner, J. E. Climate change: The cloud conundrum. *Nature* **432**, 962–963 (2004).
- Dai, A., Trenberth, K. E. & Karl, T. R. Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range. *J. Clim.* **12**, 2451–2473 (1999).
- Gómez, I. E. & Fernández, W. Variación interanual de la temperatura en Costa Rica. *Top. Meteor. Oceanogr.* **3**, 27–44 (1996).
- Quintana-Gómez, R. A. Trends of maximum and minimum temperatures in northern South America. *J. Clim.* **12**, 2104–2112 (1999).
- Cáceres, L., Mejía, R. & Otaneda, G. in *Consecuencias Climáticas e Hidrológicas del Evento El Niño a Escala Regional y Local. Incidencia en América del Sur* (eds Cadier, E., Gómez, G., Galarraga, R. & Fernández-Jáuregui, C.) (<http://www.unesco.org.uy/phi/libros/enso/caceres.html>) (UNESCO-PHI, 1998).
- Sanchez-Azofeifa, G. A., Harriss, R. C. & Skole, D. L. Deforestation in Costa Rica: a quantitative analysis using remote sensing imagery. *Biotropica* **33**, 378–384 (2001).
- Pielke, R. A. et al. A comprehensive meteorological modeling system—RAMS. *Meteorol. Atmos. Phys.* **49**, 69–91 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. P. Mitchell at JISAO and S. J. Worley at NCAR for help in obtaining climate data, R. A. Alford, F. Bolaños, J. P. Collins, R. O. Lawton, K. R. Lips, M. D. Mastrandrea, K. G. Murray, P. Ramírez and B. D. Santer for discussion, and the many contributors to the *Atelopus* database. S. H. Schneider, A. R. Blaustein and C. Parmesan commented on earlier drafts of the manuscript. The Declining Amphibian Populations Task Force and Conservation International's Critically Endangered Neotropical Species Fund provided partial funding to J.A.P. The Canada Foundation for Innovation and the Tinker Foundation helped produce the remote-sensing databases. The Research and Analysis Network for Neotropical Amphibians and the US National Science Foundation sponsored meetings that facilitated portions of the study.

Author Contributions All authors after the first are listed alphabetically. J.A.P.

conceived, designed and orchestrated the study, conducted most of the analyses (principally with J.A.C. and K.L.M.), formulated the chytrid-thermal-optimum hypothesis (with R.P.), and wrote the paper (with editing by J.A.C. and K.L.M.). M.R.B., L.A.C., M.P.L.F., E.L.M., A.M.-V. and S.R.R. provided key data and background information. E.L.M. compiled the *Atelopus* database (with B.E.Y.). P.N.F. conducted the climate simulations. G.A.S.-A. analysed the remote-sensing data. C.J.S. helped with the climate analyses and their interpretation. B.E.Y. obtained funding for meetings, provided logistics, and analysed GAA data for New World amphibians.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A.P. (goldtoad@racsa.co.cr).

ARTICLES

A quantitative protein interaction network for the ErbB receptors using protein microarrays

Richard B. Jones^{1*}, Andrew Gordus^{1,2*}, Jordan A. Krall¹ & Gavin MacBeath¹

Although epidermal growth factor receptor (EGFR; also called ErbB1) and its relatives initiate one of the most well-studied signalling networks, there is not yet a genome-wide view of even the earliest step in this pathway: recruitment of proteins to the activated receptors. Here we use protein microarrays comprising virtually every Src homology 2 (SH2) and phosphotyrosine binding (PTB) domain encoded in the human genome to measure the equilibrium dissociation constant of each domain for 61 peptides representing physiological sites of tyrosine phosphorylation on the four ErbB receptors. This involved 77,592 independent biochemical measurements and provided a quantitative protein interaction network that reveals many new interactions, including ones that fall outside of our current view of domain selectivity. By slicing through the network at different affinity thresholds, we found surprising differences between the receptors. Most notably, EGFR and ErbB2 become markedly more promiscuous as the threshold is lowered, whereas ErbB3 does not. Because EGFR and ErbB2 are overexpressed in many human cancers, our results suggest that the extent to which promiscuity changes with protein concentration may contribute to the oncogenic potential of receptor tyrosine kinases, and perhaps other signalling proteins as well.

The four human ErbB receptors induce a wide variety of cellular responses, ranging from migration to adhesion and from growth to apoptosis¹. Ligand binding to the extracellular domain promotes receptor dimerization and activation of the intracellular tyrosine kinase domain. Activated receptors phosphorylate each other on a number of tyrosine residues, which serve as docking sites for the SH2 (ref. 2) or PTB (ref. 3) domains of downstream enzymes or adaptor proteins. We asked whether a purely biophysical analysis of protein recruitment, performed on a genome-wide scale, could provide insight into the nature of ErbB signalling at a system level. Our approach, which uses microarrays of SH2 and PTB domains to measure the affinity of each domain for peptides that represent physiological sites of tyrosine phosphorylation, yielded the following insights: (1) investigating interactions in a non-competitive format reveals high-affinity binding sites for SH2 and PTB domains, many of which do not conform to consensus recognition sequences^{4,5}; (2) the recruitment sites on ErbB2 are much more promiscuous than those on the other receptors; (3) when only the highest affinity interactions are considered, the proteins that bind to EGFR constitute a small subset of those that bind to ErbB3; and (4) EGFR and ErbB2 become much more promiscuous when their concentration is raised, whereas ErbB3 does not. This, we propose, contributes to the high oncogenic potential of EGFR and ErbB2 and suggests alternative strategies for therapeutic intervention.

SH2 and PTB domains

To explore protein recruitment on a genome-wide scale, we began by cloning, expressing and purifying every SH2 and PTB domain encoded in the human genome. From an initial list of 109 SH2 domains and 44 PTB domains, we were able to obtain

sequence-verified clones for 106 SH2 domains and 41 PTB domains (Supplementary Table 1 and Supplementary Fig. 1). Some human proteins contain two SH2 or PTB domains. When these tandem domains are close together, their adjacency can affect their recognition properties⁶. We therefore cloned not only the isolated domains, but also the ten tandem SH2 domains and three tandem PTB domains found in the human genome.

Because our goal was to generate high-quality, quantitative information, we chose to purify the domains from large-scale bacterial cultures, rather than to use high-throughput methods. After isolating each domain, we assessed its purity by electrophoresis (Supplementary Fig. 2) and its aggregation state by gel filtration (Supplementary Table 2). Soluble protein was obtained for 140 of the 160 constructs. All but one of the remaining constructs were produced as insoluble protein, purified under denaturing conditions, and subsequently refolded (Supplementary Table 1). Notably, 13 of the 14 SH2 domains derived from the STAT and SOCS families of proteins required refolding, and eight did not contain any monomeric protein. It is likely that any interactions observed with these aggregated domains are nonspecific in nature.

ErbB peptides

In order to focus on physiologically relevant interactions, we searched the literature for experimentally verified sites of tyrosine phosphorylation on the ErbB receptors. We uncovered 12 sites on EGFR, 6 on ErbB2 and 11 on ErbB3. No experimentally verified sites were found on ErbB4. We elected, however, to include four tyrosines that were predicted by similarity to be sites of autophosphorylation (<http://us.expasy.org/sprot/>). As surrogates for the activated receptors, we synthesized 17–19-residue, phosphotyrosine

¹Department of Chemistry and Chemical Biology, and ²Program in Biophysics, Harvard University, Cambridge, Massachusetts 02138, USA.

*These authors contributed equally to this work.

(pY)-containing peptides (Supplementary Table 3). In four instances, two sites of phosphorylation lie within two residues of each other, prompting us to synthesize doubly phosphorylated peptides in addition to the singly phosphorylated ones. We also prepared non-phosphorylated versions of each peptide to serve as controls. A fluorescent dye was appended to the amino terminus of each peptide to visualize binding.

Protein microarrays

While it is a daunting task to study the interaction of 159 proteins with 66 peptides, protein microarray technology facilitates such large-scale analyses^{7,8}. We set out to fabricate microarrays of the purified domains that could then be queried with each fluorescent peptide. To reduce experimental variation and to expedite the processing of hundreds of arrays, we developed a strategy to produce protein microarrays in microtitre plates (see Methods). Samples were printed in duplicate and a small amount of cyanine-5 (Cy5)-labelled albumin was introduced into each sample to facilitate image analysis (Fig. 1a).

Although probing a protein array with a molecule of interest identifies a subset of interactions, the resulting information can be misleading. We have learned not to rely on data produced using a single concentration of a solution-phase probe, nor to rely on the intensity of individual spots, which does not always reflect the strength of the interaction. To circumvent these limitations, protein arrays were probed with eight concentrations of each peptide, ranging from 10 nM to 5 μ M (Fig. 1b), yielding saturation binding curves for each peptide–protein pair. Under equilibrium conditions, the mean fluorescence of duplicate spots (F_{obs}) can be described by equation (1),

$$F_{\text{obs}} = \frac{F_{\text{max}}[pep]}{K_D + [pep]} \quad (1)$$

where F_{max} is the maximum fluorescence at saturation, $[pep]$ is the total peptide concentration, and K_D is the apparent equilibrium dissociation constant. For each peptide, we fit all 159 curves, one for each single or tandem domain. Because nonspecific binding increases linearly with peptide concentration, whereas specific binding saturates, we scored as ‘specific’ those interactions that fit well to equation (1) ($R^2 \geq 0.9$), with a K_D below 2 μ M and an F_{max} at least twofold higher than the mean fluorescence of control spots. The curves that met these criteria for ErbB2 pY1139 are shown in Fig. 1c. We also identified weaker interactions, which we recorded as ‘> 2 μ M’. Following this strategy, we performed the quantitative analysis illustrated in Fig. 1 for all 66 peptides. Five of the peptides exhibited high levels of background binding (Supplementary Table 3), precluding an analysis of their interactions. High quality data were obtained for the other 61 peptides (Supplementary Table 4). These data are also available on our website (<http://www.cgr.harvard.edu/macbeath/data/data.html>) and will be updated as additional sites are identified. Of the 5,247 interactions that we measured with pY-containing peptides, 353 (6.7%) exhibit K_D values below 2 μ M. If we consider ‘active’ a domain that recognizes one or more phosphopeptides, we found that at least 102 of the 115 SH2-containing constructs (89%) and 27 of the 44 PTB-containing constructs (61%) are active on the arrays. It is likely many of the ‘inactive’ PTB domains are functional, but their role is not to bind to sites of tyrosine phosphorylation⁹.

To assess the accuracy of our method, we measured K_D values for eight domain–peptide pairs using surface plasmon resonance (SPR). In each case, the free energy of binding (ΔG) calculated from our microarray experiments matched within 5% the value measured by SPR (Fig. 1d).

Quantitative protein interaction networks

Using the data derived from this large-scale analysis, we constructed a

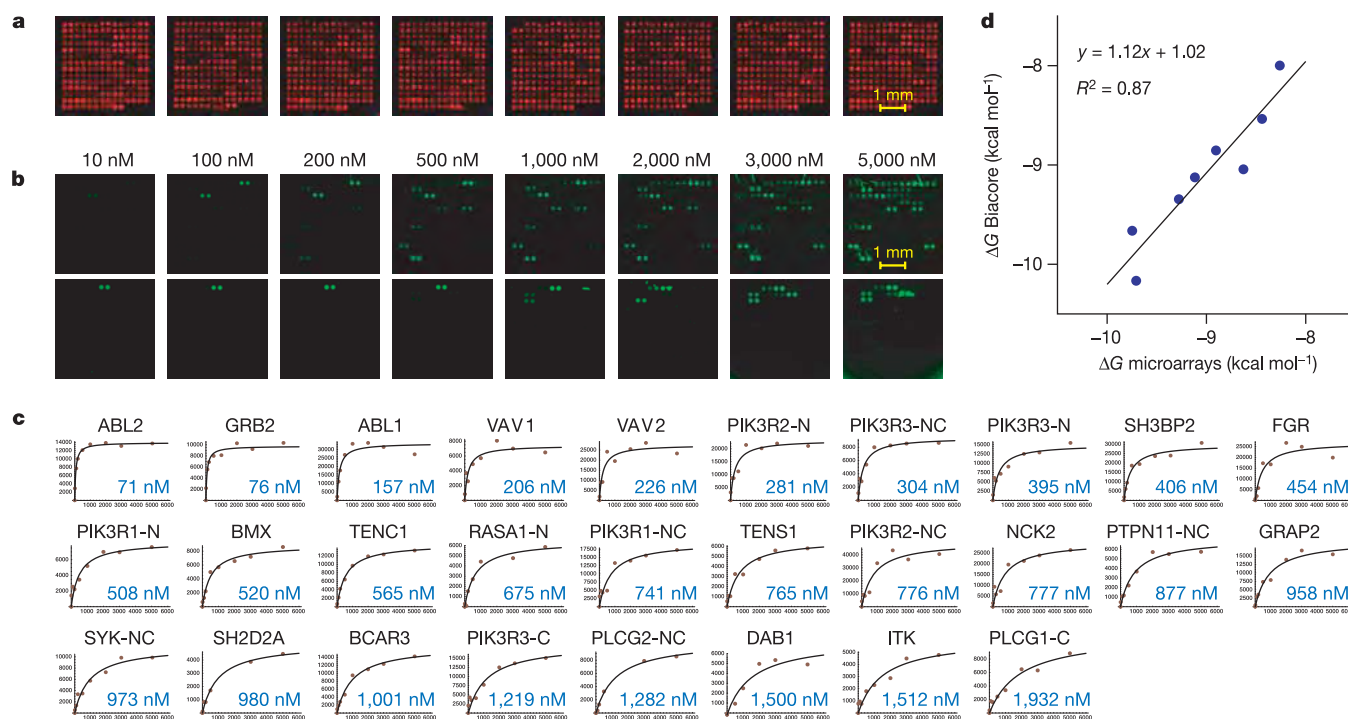


Figure 1 | Measuring the binding affinity of SH2/PTB domains for phosphopeptides derived from the ErbB receptors using protein microarrays. **a**, Fluorescent images of eight identical SH2/PTB microarrays in separate wells of a 96-well microtitre plate. The fluorescence arises from a trace amount of Cy5-labelled BSA that was added to each protein before arraying. **b**, Fluorescent images of SH2/PTB microarrays, probed with eight

different concentrations of a 5(6)-TAMRA-labelled phosphopeptide derived from ErbB2 (pY1139). **c**, Plots showing fluorescence as a function of peptide concentration for 28 high-affinity interactions. The data were fit to equation (1) to determine the apparent K_D . **d**, Comparison of the free energy of binding for eight domain–peptide interactions measured using protein microarrays with those measured using SPR (Biacore).

graphical representation of pY-mediated recruitment (Fig. 2). These diagrams provide a system-level view of the ErbB receptors, showing biophysical interactions between signalling proteins and known sites of tyrosine phosphorylation. Which proteins are actually recruited in a given cell will depend on many factors, including the effective concentrations of both the activated receptors and the signalling proteins. These diagrams should therefore be viewed as quantitative maps of the receptors, rather than a depiction of protein recruitment in any specific cell type or state.

To evaluate how well our microarray experiments recapitulate known interactions, we compiled a list of previously reported interactions between SH2/PTB-containing proteins and the ErbB receptors (Supplementary Table 5). For interactions with EGFR

and ErbB2, we relied on hand-curated databases (ref. 10 and <http://proteome.incyte.com/>); for ErbB3 and ErbB4, we surveyed the literature ourselves. Overall, our arrays detected 43 of the 65 previously reported interactions. For example, we observed that peptides derived from EGFR were able to bind strongly ($K_D < 2 \mu\text{M}$) to the SH2/PTB domains of Crk, Grb2, Nck1, PI3K α (also known as PIK3R1), PI3K β (also known as PIK3R2), PLC- γ 1 (also known as PLCG1), PLC- γ 2 (also known as PLCG2), Shp2 (also known as PTPN11), RasGAP (also known as RASA1), Shc1, Shc3, Syk and Vav1, and weakly to the SH2 domains of Grb10, Grb7, Nck2, Shp1 (also known as PTPN6), Nsp1 (also known as SH2D3A), Socs1, Stat1, Stat3, Vav2 and Vav3. Many of the known interactions that were not detected were members of the STAT and SOCS families of

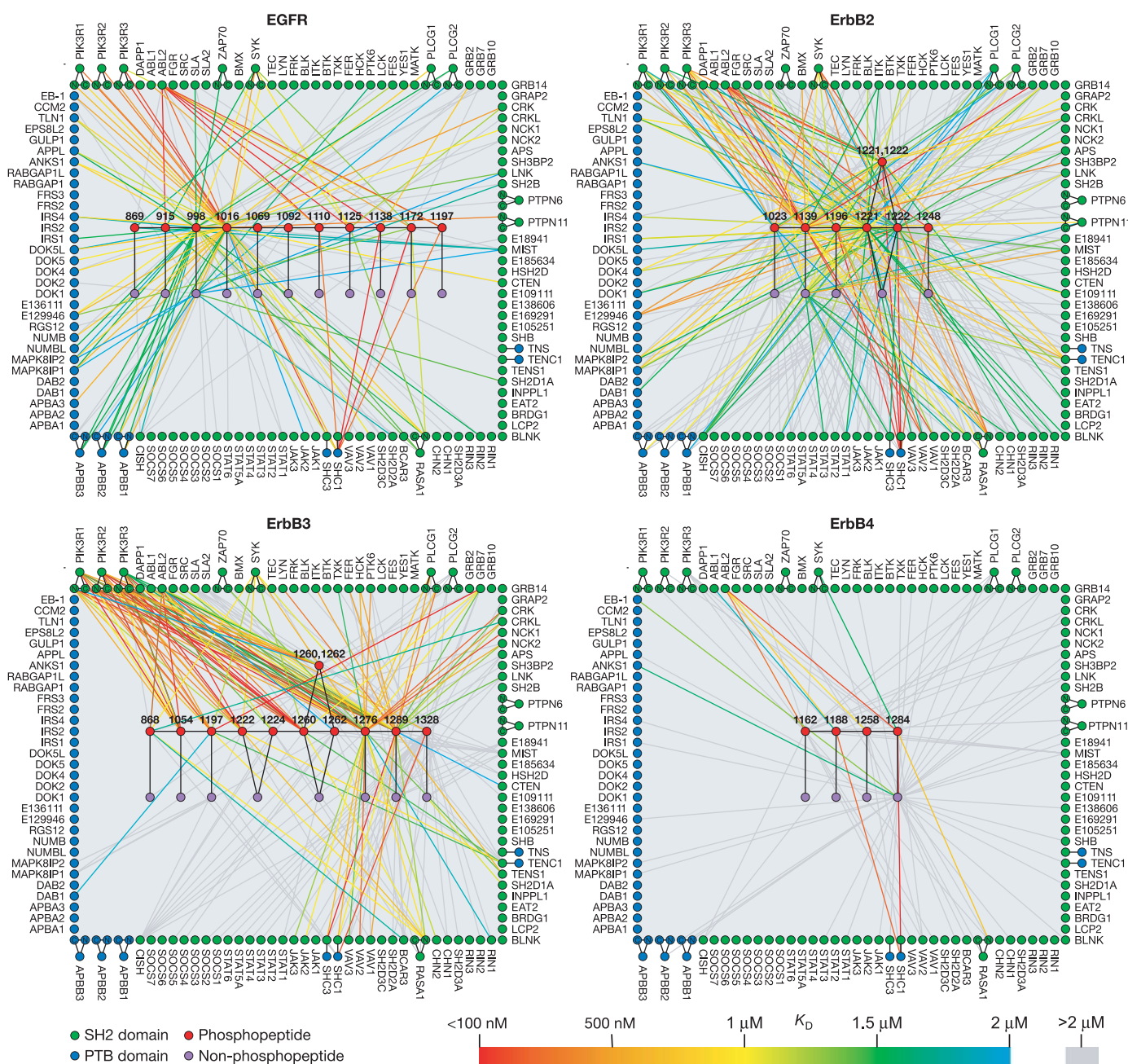


Figure 2 | Quantitative protein interaction networks for the four human ErbB receptors. Red circles represent phosphopeptides; purple circles represent the non-phosphorylated version of each phosphopeptide; green circles represent SH2 domains; and blue circles represent PTB domains. Lines connecting peptides to domains indicate observed interactions,

coloured according to the affinity of the interaction (see legend). Red circles labelled with two numbers represent doubly phosphorylated peptides. The green or blue circles that lie outside the rectangle of individual domains represent tandem domains. Black lines connect the tandem domains to their corresponding individual domains.

proteins that contained a high percentage of aggregated protein. In addition, because many reported interactions are based on co-purification experiments, some may not depend on SH2/PTB domains or may be mediated by bridging proteins.

We also compared our observed interactions with those predicted by Scansite 2.0 (refs 11, 12), a program that uses the consensus binding information provided by oriented-peptide library screens^{4,5} to predict interactions between domains and sequences of interest. We scanned the 14 SH2 domains and one PTB domain available in Scansite against the peptide sequences used in our study. For comparison purposes, we considered only the 11 domains that bind more than one peptide on the arrays and are predicted by Scansite to recognize more than one peptide in our study. When run on its lowest stringency setting, Scansite predicts 56% of the strong ($K_D < 2 \mu\text{M}$) interactions that we observe, and 47% of all interactions, indicating that many peptides with sequences that do not conform to consensus motifs are recognized by these domains. For example, Scansite does not predict an interaction between the SH2 domain of Abl1 and ErbB2 pY1139, yet the arrays show a strong interaction ($K_D = 157 \text{ nM}$), which we also confirm by SPR ($K_D = 140 \text{ nM}$). Conversely, 51% of the Scansite hits are not observed on the arrays. Overall, we find that Scansite predictions closely match the microarray data for the PTB domain of Shc1 and the SH2 domains of PI3K and Grb2, but the overlap is much less for other domains (Fig. 3). We attribute this difference to the non-competitive format of the microarray experiment in which only physiologically relevant sequences are assessed without interference from tight-binding but irrelevant peptides.

Not surprisingly, the SH2/PTB microarrays uncovered many strong interactions ($K_D < 2 \mu\text{M}$) that have not been reported previously: 32 with EGFR, 48 with ErbB2, 33 with ErbB3 and 3 with ErbB4 (Supplementary Table 5). For example, the arrays revealed that the SH2 domain of v-crk avian sarcoma virus CT10-homologue-like protein (CrkL) recognizes phosphopeptides derived from ErbB2 and ErbB3. Consistent with this observation, we found that CrkL becomes phosphorylated on Y207 in A431 squamous carcinoma

cells and in MDA-MB-468 breast carcinoma cells within 1 min of stimulation with EGF, and in MDA-MB-468 and T-47D breast carcinoma cells within 5 min of stimulation with heregulin- $\beta 1$ (HRG- $\beta 1$, a ligand for ErbB3; Supplementary Fig. 3a, b). The slower phosphorylation of CrkL in response to HRG- $\beta 1$ stimulation is probably due to the substantially lower number of ErbB3 molecules in MDA-MB-468 and T-47D cells relative to the number of EGFR and ErbB2 molecules in A431 and MDA-MB-468 cells¹³.

System-level properties of ErbB receptors

Whereas our arrays provide a list of previously unrecognized biochemical interactions, as well as data that can fuel efforts to build computational models of signal transduction¹⁴, they also offer an unbiased, system-level view of the ErbB receptors that was previously unavailable. The networks of Fig. 2 show that EGFR and ErbB3 each have two sites (Y998 and Y1016 on EGFR, and Y1276 and Y1289 on ErbB3) that can engage in many high-affinity interactions. These tyrosines may serve as 'multifunctional docking sites'¹⁵ that have different roles depending on the relative concentrations of their target proteins. The other phosphotyrosines on these receptors are markedly more selective and presumably serve specialized functions.

In contrast, ErbB2 features many highly promiscuous sites. When considering only strong interactions ($K_D < 2 \mu\text{M}$), the sites on ErbB2 bind over 17 different proteins on average, whereas those on EGFR, ErbB3 and ErbB4 bind 7.2, 8.8 and 2.3 proteins on average, respectively. Unlike the other receptors, ErbB2 does not recognize an extracellular ligand, but instead functions primarily as a heterodimerization partner. It has been shown that ErbB2 quantitatively increases both the amplitude and duration of EGFR signalling by increasing the ratio of active kinase to ligand and by inhibiting the downregulation of EGFR¹⁶. Our data suggest that ErbB2 may also qualitatively expand the diversity of signalling by recruiting proteins that the other receptors cannot. The sparse connections to ErbB4 may indicate that this receptor serves a more specialized function than the others. It is likely, however, that we are missing sites of tyrosine phosphorylation on ErbB4, and that even these four sites may not be physiologically relevant.

The networks of Fig. 2 show that, in principle, some phosphotyrosines can recruit many different proteins. For example, pY1139 of ErbB2 interacts strongly ($K_D < 2 \mu\text{M}$) with 24 different proteins. Although SH2/PTB-containing proteins are produced at different levels in different cells, many that are able to bind to the same site are co-expressed¹⁷. Given the high connectivity of the network, it is possible that some sites serve more than one function within the same cell at the same time. The six tightest binders of ErbB2 pY1139 have K_D values that fall within less than fourfold of each other. If two of these proteins are present in the vicinity of the activated receptor at similar concentrations, both may be recruited and activated.

It is tempting to speculate that some sites perform different functions based on the location of the receptor in the cell. All of the interactions that we have investigated by SPR display half-lives of less than 10 s, which indicates that interactions based solely on SH2 domains are highly dynamic¹⁸. When a receptor is initially activated, proteins that function at the cell surface (such as PI3K and PLC- γ) are recruited. As the receptor is internalized, the repertoire of proteins it encounters changes, leading to the recruitment of a different set of proteins. A single site could thus serve more than one function in the same cell, at the same time, but in different locations. It is now clear that internalized receptors continue to signal¹⁹ and that some proteins preferentially associate with surface receptors, whereas others bind almost exclusively to intracellular receptors²⁰. The protein recruitment profile of a cell is therefore likely to depend not only on which proteins are co-expressed, but on receptor trafficking and partitioning as well.

Networks at different affinity thresholds

Most interaction networks reported to date are boolean: proteins

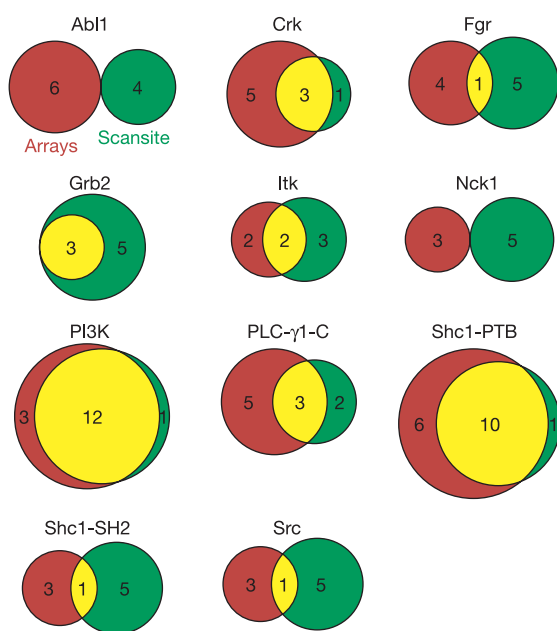


Figure 3 | Venn diagrams for 11 of the 15 SH2/PTB domains available in Scansite 2.0. For each domain, the red circle represents the phosphopeptides that are observed to bind to that domain using protein microarrays and the green circle represents the phosphopeptides that are predicted to bind to that domain by Scansite 2.0 (run on its lowest stringency setting).

either 'interact' or 'don't interact'^{21–24}. Such networks represent a single slice through interaction space, where the slice is made at an affinity threshold defined by the assay. Because we quantified every interaction, we can view different slices through the network at defined affinity thresholds. To a first approximation, these slices highlight interactions that are relevant at different concentrations of activated receptor. At low levels of receptor activation, or in cells in which the receptor is expressed at low levels, only the highest affinity interactions are important. At higher levels of activated receptor, however, the lower affinity interactions are also relevant (assuming the receptor is present in excess). Although the protein recruitment profile varies from one cell to the next based on affinities and concentrations, the general principles that arise from this analysis provide insight into the intrinsic properties of each receptor. Figure 4 shows four slices through the ErbB network at different affinity thresholds. ErbB4 was excluded due to the paucity of data. EGFR, ErbB2 and ErbB3, however, exhibit intriguing differences. Notably, the ErbB3 network changes very little as the threshold is lowered, whereas EGFR and ErbB2 become much more promiscuous (Fig. 4a, b). The biological implication of this finding is that cells should be less sensitive to changes in the levels of ErbB3 relative to EGFR and ErbB2. Consistent with this prediction, EGFR and ErbB2 expression levels vary more across normal human tissues than do those of ErbB3 (ref. 17).

The oncological implication is equally intriguing. The threshold slices of Fig. 4 suggest that elevated levels of ErbB3 should primarily induce stronger signalling through pathways that are normally activated by low levels of ErbB3. In contrast, elevated levels of EGFR or ErbB2 should induce signalling through alternative pathways that are not activated at lower levels. Interestingly, overexpression, gene amplification, or overactivation of EGFR and ErbB2 are frequently observed in human cancers^{25–30}, whereas there is no evidence for gene amplification of ErbB3, and overexpression is limited¹. We propose that the high oncogenic potential of EGFR and ErbB2 arises, at least in part, from their ability to turn on different pathways when overexpressed, rather than simply from stronger signalling through their primary pathways. Secondary signalling proteins that are only recruited by overexpressed receptors may also engage in cross-talk with primary signalling proteins, further altering the state of the cell. This hypothesis suggests that proteins involved in the most critical secondary pathways may serve as more selective targets for cancer chemotherapy.

We also asked what each receptor can do that the others cannot, and how this changes with receptor levels. Figure 4c shows Venn diagrams indicating the number of proteins that can be recruited to each receptor at each affinity threshold. When only the tightest interactions are considered ($K_D < 500$ nM), there is only one protein that binds solely to EGFR (Shp2/PTPN11). With this one exception,

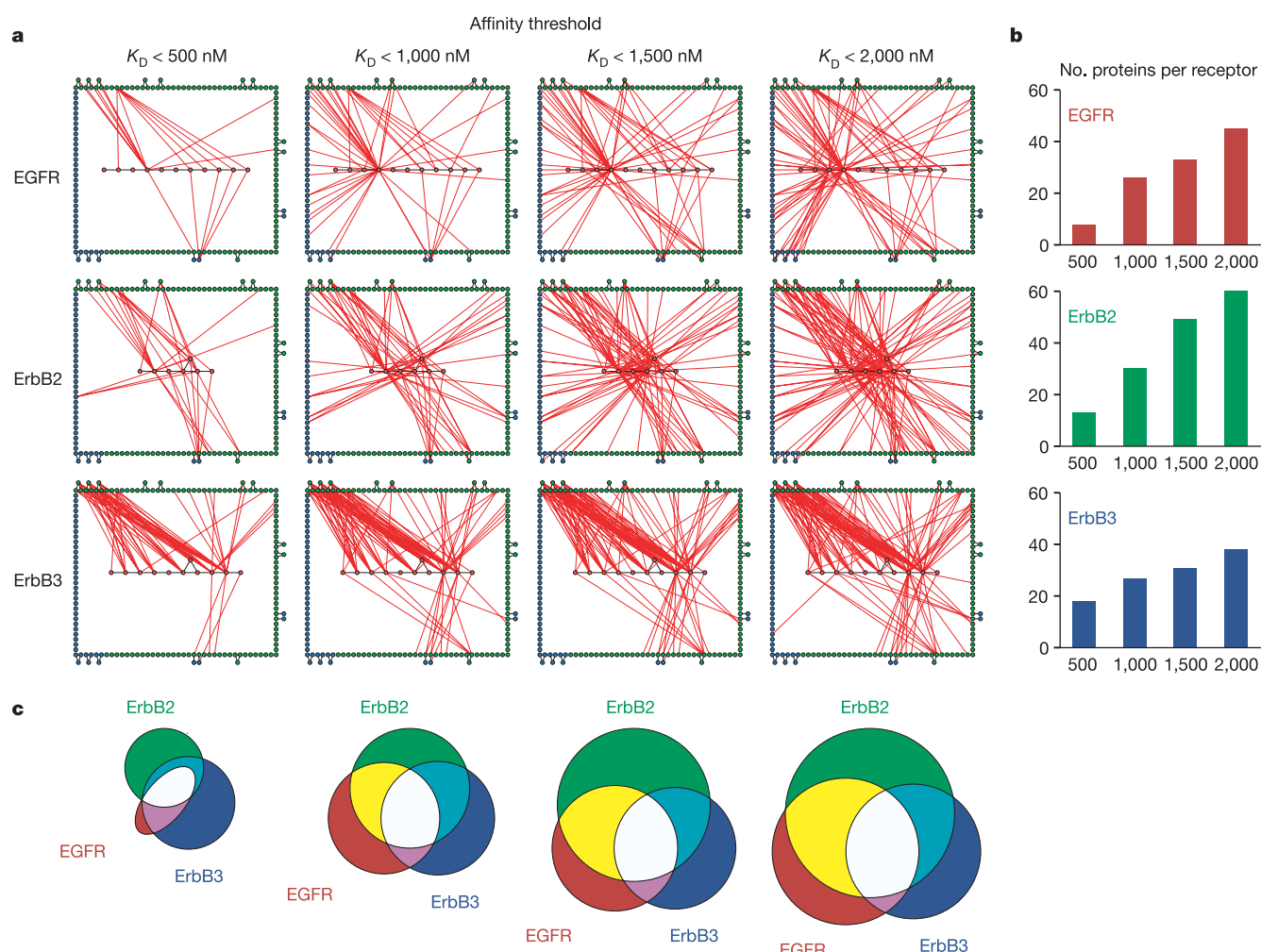


Figure 4 | A system-level view of EGFR, ErbB2 and ErbB3 at different affinity thresholds. **a**, Four views of the ErbB interaction networks, each at a different affinity threshold. At each threshold, only interactions with a K_D below the indicated value are shown. **b**, Plots illustrating the number of proteins that bind to each receptor at each affinity threshold. The y axis

shows the number of different proteins that contain at least one SH2 or PTB domain with a K_D value below that indicated on the x axis (nM). **c**, Venn diagrams, drawn to scale, illustrating the number of proteins that are recruited to each receptor at each affinity threshold.

the proteins that bind to EGFR constitute a small subset of those that bind to ErbB3. It is only at high receptor levels that EGFR is able to recruit proteins that ErbB2 and ErbB3 cannot. Even at low concentration, on the other hand, ErbB2 recruits proteins that ErbB3 cannot. Our results indicate that, under stringent conditions, the ErbB2–ErbB3 complex is broadest in scope, followed by EGFR–ErbB3, EGFR–ErbB2 and finally EGFR–EGFR. Consistent with this observation, ErbB2–ErbB3 is the most transforming receptor complex^{31,32} and the mitogenicity of these homo- and heterodimers tracks perfectly with their breadth of signalling³³.

What proteins are common to all three receptors? In general, they are the ones that initiate canonical signalling pathways: PI3K, which regulates cell cycle progression, cell growth, cytoskeletal rearrangements and vesicular transport; Shc1, Crk and RasGAP, which feed into the MAPK cascade to control proliferation and motility; and Syk and PLC- γ , which regulate cytoskeleton–membrane interactions. We were surprised, however, to find that Abl1 and Abl2 bind with high affinity to sites on all four receptors, even though they are not typically considered in the context of ErbB signalling. Nevertheless, we found that Abl1 is phosphorylated on Y412 in A431 cells within 1 min of treatment with EGF (Supplementary Fig. 3a, c). Abl1 normally controls cytoskeletal rearrangements³⁴ and thus may regulate cell migration or adhesion in response to EGF and its relatives.

Discussion

By performing a comprehensive analysis of SH2/PTB-mediated interactions with the ErbB receptors, we uncovered a quantitative network that reveals the ability of each receptor to recruit signalling proteins upon activation. In addition to confirming 43 previously recognized interactions, we identified 116 new biophysical interactions and provided evidence for the physiological relevance of several of them. Most importantly, the network defined by this effort provides system-level insight into ErbB function and reveals a surprising property of receptor tyrosine kinases: they differ in the extent to which they become more promiscuous when overexpressed. This observation highlights the importance of collecting quantitative information on protein–protein interactions. It is our hope that modelling studies based on a marriage of our biophysical measurements with quantitative cell biological data will prove useful in predicting normal cellular behaviour, as well as how best to intervene when signalling goes awry.

METHODS

Cloning of SH2 and PTB domains. When we began this effort, a search of three databases—SMART³⁵, Pfam³⁶ and Ensembl (<http://www.ensembl.org/>)—yielded a list of 109 SH2 domains, 44 PTB domains, 10 tandem SH2 domains and 3 tandem PTB domains. Sequences encoding these domains were amplified from human cDNA and transferred into pENTR/D-TOPO (Invitrogen) by topoisomerase I-mediated directional cloning. Each clone was verified by DNA sequencing.

Production and purification of recombinant proteins. SH2/PTB coding sequences were transferred into a Gateway-compatible *Escherichia coli* expression vector (pET-32-DEST). Recombinant proteins were produced in 500-ml cultures and purified by immobilized metal affinity chromatography. Purified proteins were dialysed against buffer A (300 mM NaCl, 50 mM Na₂PO₄, pH 8) and glycerol was added to a final concentration of 20% (v/v).

Peptide synthesis. Peptides were synthesized on the solid phase (50 μ mol scale) using Fmoc chemistry. Peptides were labelled on their amino termini with 5- and 6-carboxytetramethylrhodamine (5(6)-TAMRA) before deprotection and cleavage. All peptides were purified by reverse phase high-performance liquid chromatography (HPLC).

Fabrication and processing of protein microarrays. Purified SH2/PTB domains were spotted in duplicate at a concentration of 40 μ M onto aldehyde-modified glass substrates (112.5 mm $\times$ 74.5 mm $\times$ 1 mm) using a piezoelectric microarrayer. Ninety-six identical arrays were fabricated in a 12 $\times$ 8 pattern to match the spacing of a microtitre plate. Each array consisted of a 14 $\times$ 14 pattern of spots, with a 250 μ m pitch. After a 1 h incubation, the glass was attached to a bottomless 96-well plate using an intervening silicone gasket. Immediately

before use, the plates were quenched with buffer B (20 mM HEPES, 100 mM KCl, 0.1% Tween-20, pH 7.8) containing 1% BSA (w/v). Arrays were probed with 5(6)-TAMRA-labelled peptides, dissolved in buffer B. After a 30-min incubation, the arrays were washed with buffer B, rinsed with ddH₂O, and spun upside down to remove residual water.

Scanning and analysis of microarrays. Protein microarrays were scanned at 10- μ m resolution using an LS400 scanner (Tecan). Spots were defined using the Cy-5 image and the mean fluorescence of each spot was calculated from the 5(6)-TAMRA image. Concentration-dependent measurements were fit to equation (1) for each domain–peptide pair and the resulting data were displayed graphically using Cytoscape 2.1 (<http://www.cytoscape.org/>).

Surface plasmon resonance. SPR studies were performed using a Biacore 3000. SH2 domains were produced as glutathione S-transferase (GST) fusion proteins and captured on the chip by an immobilized anti-GST antibody. Peptides were introduced in the solution phase and binding was measured under equilibrium conditions.

Cell culture and immunoblots. Cells were grown to approximately 70% confluence, serum-starved for 24 h, treated with either 100 ng ml⁻¹ EGF or 10 nM heregulin- β 1, and analysed by immunoblotting. Technical details of all experiments are provided in the Supplementary Methods.

Received 7 July; accepted 30 August 2005.

Published online 6 November 2005.

- Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signalling network. *Nature Rev. Mol. Cell Biol.* **2**, 127–137 (2001).
- Sadowski, I., Stone, J. C. & Pawson, T. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol. Cell. Biol.* **6**, 4396–4408 (1986).
- Kavanaugh, W. M. & Williams, L. T. An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. *Science* **266**, 1862–1865 (1994).
- Songyang, Z. *et al.* SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, 767–778 (1993).
- Songyang, Z. *et al.* Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Mol. Cell. Biol.* **14**, 2777–2785 (1994).
- Hatada, M. H. *et al.* Molecular basis for interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor. *Nature* **377**, 32–38 (1995).
- MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760–1763 (2000).
- Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
- Yan, K. S., Kuti, M. & Zhou, M. M. PTB or not PTB—that is the question. *FEBS Lett.* **513**, 67–70 (2002).
- Peri, S. *et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* **13**, 2363–2371 (2003).
- Yaffe, M. B. *et al.* A motif-based profile scanning approach for genome-wide prediction of signalling pathways. *Nature Biotechnol.* **19**, 348–353 (2001).
- Obenauer, J. C., Cantley, L. C. & Yaffe, M. B. Scansite 2.0: Proteome-wide prediction of cell signalling interactions using short sequence motifs. *Nucleic Acids Res.* **31**, 3635–3641 (2003).
- Bowers, G. *et al.* The relative role of ErbB1–4 receptor tyrosine kinases in radiation signal transduction responses of human carcinoma cells. *Oncogene* **20**, 1388–1397 (2001).
- Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D. & Muller, G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Biotechnol.* **20**, 370–375 (2002).
- Ponzetto, C. *et al.* A multifunctional docking site mediates signalling and transformation by the hepatocyte growth factor/scatter factor receptor family. *Cell* **77**, 261–271 (1994).
- Hendriks, B. S., Orr, G., Wells, A., Wiley, H. S. & Lauffenburger, D. A. Parsing ERK activation reveals quantitatively equivalent contributions from epidermal growth factor receptor and HER2 in human mammary epithelial cells. *J. Biol. Chem.* **280**, 6157–6169 (2005).
- Su, A. I. *et al.* Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl Acad. Sci. USA* **99**, 4465–4470 (2002).
- Felder, S. *et al.* SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange. *Mol. Cell. Biol.* **13**, 1449–1455 (1993).
- Haugh, J. M. Localization of receptor-mediated signal transduction pathways: the inside story. *Mol. Interv.* **2**, 292–307 (2002).
- Burke, P., Schooler, K. & Wiley, H. S. Regulation of epidermal growth factor receptor signalling by endocytosis and intracellular trafficking. *Mol. Biol. Cell* **12**, 1897–1910 (2001).
- Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).

23. Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
24. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
25. Gorgoulis, V. *et al.* Expression of EGF, TGF- α and EGFR in squamous cell lung carcinomas. *Anticancer Res.* **12**, 1183–1187 (1992).
26. Irish, J. C. & Bernstein, A. Oncogenes in head and neck cancer. *Laryngoscope* **103**, 42–52 (1993).
27. Moscatello, D. K. *et al.* Frequent expression of a mutant epidermal growth factor receptor in multiple human tumors. *Cancer Res.* **55**, 5536–5539 (1995).
28. Ross, J. S. & Fletcher, J. A. The HER-2/neu oncogene in breast cancer: prognostic factor, predictive factor, and target for therapy. *Stem Cells* **16**, 413–428 (1998).
29. Slamon, D. J. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177–182 (1987).
30. Wong, A. J. *et al.* Structural alterations of the epidermal growth factor receptor gene in human gliomas. *Proc. Natl Acad. Sci. USA* **89**, 2965–2969 (1992).
31. Alimandi, M. *et al.* Cooperative signalling of ErbB3 and ErbB2 in neoplastic transformation and human mammary carcinomas. *Oncogene* **10**, 1813–1821 (1995).
32. Wallasch, C. *et al.* Heregulin-dependent regulation of HER2/neu oncogenic signalling by heterodimerization with HER3. *EMBO J.* **14**, 4267–4275 (1995).
33. Pinkas-Kramarski, R. *et al.* Diversification of Neu differentiation factor and epidermal growth factor signalling by combinatorial receptor interactions. *EMBO J.* **15**, 2452–2467 (1996).
34. Woodring, P. J., Hunter, T. & Wang, J. Y. Regulation of F-actin-dependent processes by the Abl family of tyrosine kinases. *J. Cell Sci.* **116**, 2613–2626 (2003).
35. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: identification of signalling domains. *Proc. Natl Acad. Sci. USA* **95**, 5857–5864 (1998).
36. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **28**, 263–266 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Schoeberl for comments on the manuscript, J. Grudzien for development work on the silicone gaskets, and the Bauer Center for Genomics Research at Harvard University for support with instrumentation and automation. This work was supported by awards from the W. M. Keck Foundation and the Arnold and Mabel Beckman Foundation. R.B.J. is the recipient of a Ruth L. Kirschstein National Research Service Award (NIH), A.G. is the recipient of an NSF Graduate Research Fellowship, and J.A.K. is the recipient of a Howard Hughes Medical Institute Predoctoral Fellowship in the Biological Sciences.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.M. (macbeath@chemistry.harvard.edu).

A magnetic reconnection X-line extending more than 390 Earth radii in the solar wind

T. D. Phan¹, J. T. Gosling², M. S. Davis¹, R. M. Skoug³, M. Øieroset¹, R. P. Lin¹, R. P. Lepping⁴, D. J. McComas⁵, C. W. Smith⁶, H. Reme⁷ & A. Balogh⁸

Magnetic reconnection in a current sheet converts magnetic energy into particle energy, a process that is important in many laboratory¹, space^{2,3} and astrophysical contexts^{4–6}. It is not known at present whether reconnection is fundamentally a process that can occur over an extended region in space or whether it is patchy and unpredictable in nature⁷. Frequent reports of small-scale flux ropes and flow channels associated with reconnection^{8–13} in the Earth's magnetosphere raise the possibility that reconnection is intrinsically patchy, with each reconnection X-line (the line along which oppositely directed magnetic field lines reconnect) extending at most a few Earth radii (R_E), even though the associated current sheets span many tens or hundreds of R_E . Here we report three-spacecraft observations of accelerated flow associated with reconnection in a current sheet embedded in the solar wind flow, where the reconnection X-line extended at least $390R_E$ (or 2.5×10^6 km). Observations of this and 27 similar events imply that reconnection is fundamentally a large-scale process. Patchy reconnection observed in the Earth's magnetosphere is therefore likely to be a geophysical effect associated with fluctuating boundary conditions, rather than a fundamental property of reconnection. Our observations also reveal, surprisingly, that reconnection can operate in a quasi-steady-state manner even when undriven by the external flow.

Until recently, *in situ* observations of reconnection in space plasmas were made almost exclusively in the Earth's magnetosphere, in current sheets formed by the interaction between the solar wind and the geomagnetic field. Such current sheets have finite extents, and their boundary conditions (determined by the solar wind magnetic field) often change rapidly. It is generally difficult to establish the presence of an extended reconnection X-line in the magnetosphere from *in situ* measurements since that requires the presence of widely separated spacecraft detecting the same reconnection events. The chances of such conjunctions are exceedingly small because the spacecraft are seldom ideally positioned for such observations and because of the variable boundary conditions. The single event reported where two spacecraft (separated by $3R_E$) detected the same reconnection event at the magnetopause only allowed the deduction that the X-line was at least $3R_E$ long¹⁴. Remote observations of proton auroras¹⁵ and ionospheric convection¹⁶ have hinted at the presence of a magnetopause X-line up to $40R_E$ in length but that has not yet been confirmed by *in situ* observations.

The recent discovery of reconnection exhausts in the solar wind^{17,18} introduces a new laboratory where reconnection can be investigated by *in situ* measurements. The solar wind reconnection events are often associated with interplanetary coronal mass ejections, and the magnetic field orientations on the two sides of the current sheets are

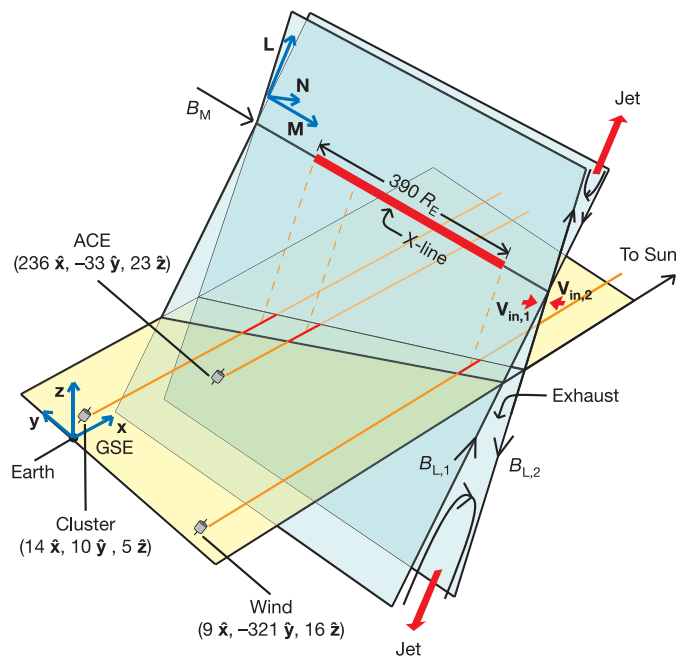


Figure 1 | Diagram of the encounters of three spacecraft with an extended ($390R_E$) magnetic reconnection X-line in the solar wind. Reconnection in the current sheet (in blue) occurs at the X-line between magnetic field lines with large anti-parallel components $B_{L,1}$ and $B_{L,2}$; the resulting bi-directional plasma jets (confined to the reconnection exhausts) can be observed far from the X-line. The ACE, Cluster and Wind spacecraft positions are shown in units of Earth radius (R_E) and in geocentric solar ecliptic (GSE) coordinates with the x-axis pointing from Earth to Sun, the y-axis pointing towards dusk and the z-axis parallel to the ecliptic pole. All three spacecraft were relatively close to the ecliptic plane (in yellow). ACE was $222R_E$ upstream of Cluster while Wind was $331R_E$ downward of Cluster. Also shown is the LMN current sheet coordinate system, with N along the overall current sheet normal, M along the X-line direction and L along the anti-parallel magnetic field direction. The current sheet normal ($0.71\hat{x}, 0.60\hat{y}, -0.37\hat{z}$) in GSE, is tilted 45° relative to the Sun–Earth line. The X-line is oriented along $(0.47\hat{x}, -0.79\hat{y}, -0.39\hat{z})$ in GSE. The thick solid red line is the ($390R_E$) portion of the X-line whose effect is observed by the three spacecraft. The solid orange lines denote the spacecraft trajectory relative to the solar wind, with the red line portion marking the intersections of the exhaust with the spacecraft. The total reconnected magnetic flux ($= V_{in,1}B_{L,1}L_{X-line}$ or $V_{in,2}B_{L,2}L_{X-line}$) is determined by the inflow velocity, V_{in} , the strength of the anti-parallel field components, B_L , and the length of the X-line, L_{X-line} . The angle of the diverging exhausts is exaggerated for illustration. The actual calculated angle is $\sim 4^\circ$. B_M is the magnetic field along the X-line.

¹Space Sciences Laboratory, University of California, Berkeley, California 94720, USA. ²Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80303, USA. ³Los Alamos National Laboratory, MS D466, Los Alamos, New Mexico 87545, USA. ⁴Laboratory for Solar and Space Physics, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ⁵Space Science and Engineering Division, Southwest Research Institute, PO Drawer 28510, San Antonio, Texas 78228, USA. ⁶Department of Physics and Space Science Center, University of New Hampshire, Room 207 Morse Hall, 30 College Road, Durham, New Hampshire 03824, USA. ⁷Centre d'Etude Spatiale des Rayonnements, BP 4346, Toulouse 31029, France. ⁸Space and Atmospheric Physics Group, Imperial College, London SW7 2BZ, UK.

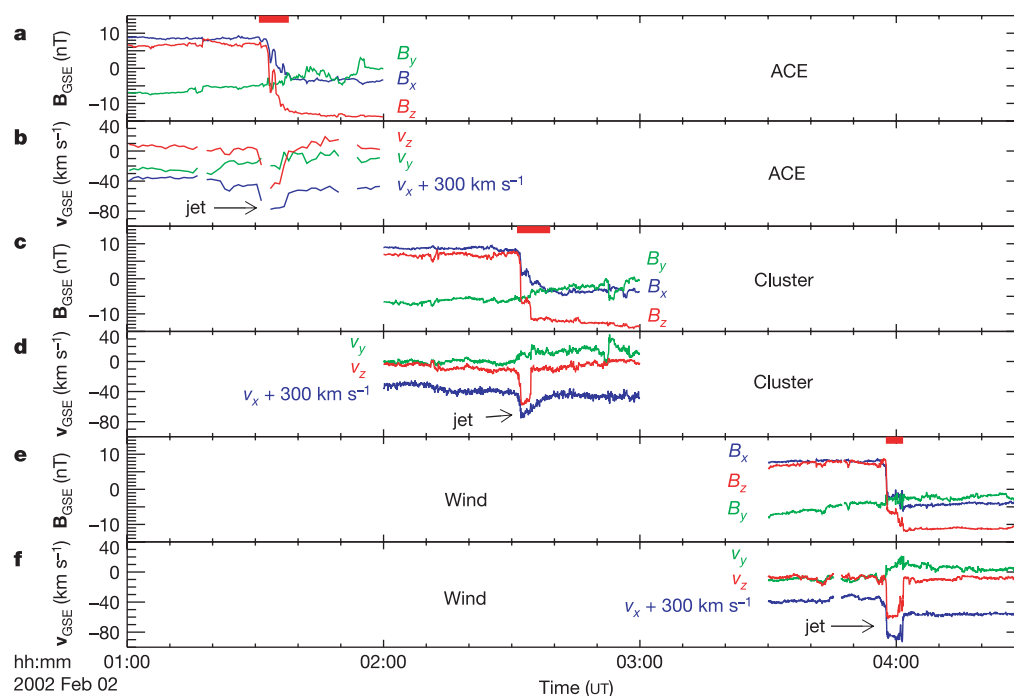


Figure 2 | Detections of the magnetic reconnection exhaust by the ACE, Cluster-3 and Wind spacecraft on 2 February 2002. **a, b,** The magnetic field and plasma velocity in GSE coordinates measured by ACE; **c, d,** the magnetic field and velocity measured by Cluster-3; and **e, f,** the magnetic field and velocity measured by Wind. The x component of the velocity in **b, d** and **f** has been shifted by $+300 \text{ km s}^{-1}$. The red horizontal bars in **a, c** and **e** indicate the durations of the encounters by the three spacecraft. The magnetic field rotated 140° across the exhaust. The plasma flow in the

exhaust was enhanced by $\sim 50 \text{ km s}^{-1}$ relative to the ambient solar wind flow speed. The velocity components were correlated (anti-correlated) with the components of the magnetic field at the leading (trailing) edge of the exhaust, as expected from reconnection sunward and northward of the spacecraft. It is concluded that all three (widely separated) spacecraft detected essentially the same reconnection signature. The abrupt changes in the magnetic field B_z at the two edges and a plateau in the B_z profile in the middle of the current sheet indicate that the current sheet is bifurcated.

usually well defined. The combination of extended current sheets with stable boundary conditions and the fact that the solar wind rapidly convects the exhausts past observing spacecraft make these solar wind reconnection events ideal for addressing the question of extended versus patchy reconnection without complications due to boundary effects.

On 2 February 2002, the Wind, ACE and Cluster spacecraft were all in the solar wind (Fig. 1). Cluster was $14R_E$ upstream (sunward) of the Earth. ACE was $222R_E$ further upstream of Cluster, while Wind, in its furthest orbit from Earth during its 10-yr mission, was located at $331R_E$ downward of Cluster (and $321R_E$ from the Sun–Earth line). Figure 2 shows that all three spacecraft detected the passage of the same bifurcated current sheet with accelerated plasma flow embedded in it. The total magnetic field rotation (or shear) across the bifurcated current sheet was 140° . The observed plasma acceleration within the exhaust agreed with the reconnection prediction to within 5° in direction and 10% in flow speed (see Fig. 3c and d for more details). This is consistent with the plasma acceleration being accomplished by the magnetic tension force associated with linkage of the magnetic field across the exhaust. Furthermore, Fig. 3 shows that the plasma density and temperature were sharply enhanced at the edges of the current sheet while the magnetic field strength was reduced. These signatures are consistent with the Petschek¹⁹ model of fast reconnection, where the reconnection exhaust is bounded by Alfvén and/or slow mode waves. The plasma and field signatures just described are typical of solar wind reconnection exhausts^{17,18}. What is significant about this 2 February 2002 event is the fact that the reconnection exhaust was observed by three widely separated spacecraft, which allows the deduction of a long reconnection X-line.

The extent of the X-line that can be measured depends on the orientation of the exhaust and of the X-line relative to the spacecraft.

To obtain the X-line orientation, one first needs to determine the exhaust geometry.

The bifurcated current sheet associated with the reconnection exhaust was convecting with the solar wind, and was first detected at ACE at $\sim 01:32 \text{ UT}$, followed by Cluster an hour later (at $\sim 02:32 \text{ UT}$) and 2.5 h later than at ACE by Wind (at $\sim 03:57 \text{ UT}$). The fact that Cluster and Wind detected the current sheet 85 min apart even though both spacecraft were at nearly the same distance from the Sun (but $330R_E$ apart in dawn–dusk direction) implies that the current sheet must make a large angle relative to the ‘east–west’ direction (that is, relative to the y direction in Geocentric Solar Ecliptic (GSE) coordinates; see Fig. 1). This angle is confirmed by the analysis of the current sheet geometry at Wind. The normal to the current sheet tilt was determined by minimum variance analysis²⁰ of the magnetic field across the current sheet, and was found to be $(0.71\hat{x}, 0.60\hat{y}, -0.37\hat{z})$ in GSE. The resulting error in the propagation time from ACE to Cluster is 4 min 20 s, or 7%. From ACE to Wind, the error is only 6 s, or 0.07%. This agreement demonstrates that the current sheet was indeed approximately flat on a scale of hundreds of Earth radii (or 0.01 AU) and that the current sheet normal was accurate. The small magnitude of the normal magnetic field (B_N) across the current sheet (Fig. 3e) further confirms the accuracy of the current sheet normal.

The X-line orientation $(0.47\hat{x}, -0.79\hat{y}, -0.39\hat{z})$ in GSE is obtained from the components of the magnetic field in the current sheet plane²¹. From the X-line orientation one can determine, based on the locations where the three spacecraft intersected the current sheet, that Cluster and Wind detected flow from positions along the X-line that were $390R_E$ apart, while ACE detected flow from the X-line at an intermediate location (see Fig. 1). This implies that the X-line extended at least $390R_E$ (or 4×10^4 ion inertial lengths) and very probably a great deal further. If reconnection were patchy, one or

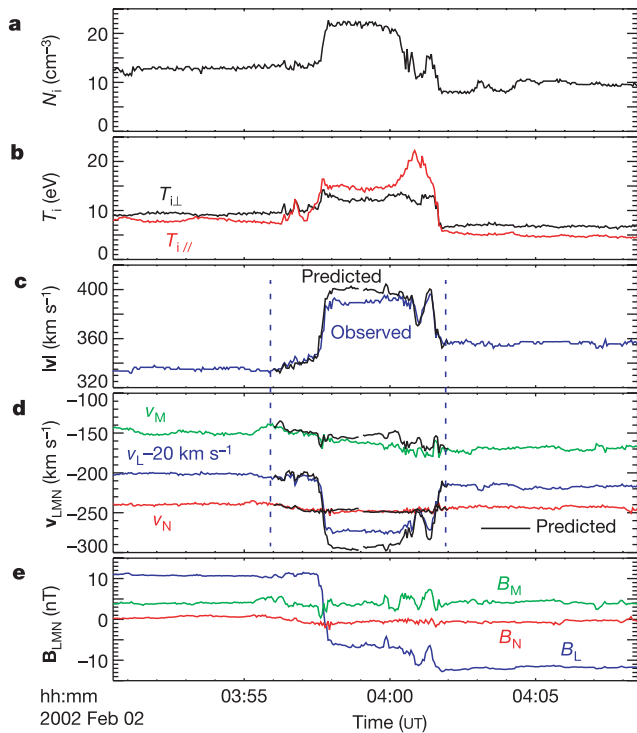


Figure 3 | Quantitative comparison between the flow acceleration observed by the Wind spacecraft and the prediction from reconnection. **a**, The ion density; **b**, the parallel and perpendicular ion temperatures; **c**, the observed and (reconnection) predicted plasma flow speed; **d**, the observed and predicted (in black) plasma velocity in LMN coordinates; and **e**, the magnetic field in LMN coordinates. The anti-parallel component of the magnetic field (B_L) was nearly equal in magnitude on the two sides of the exhaust. The guide field (along the M or X-line direction) was ~ 4 nT, or 35% of the anti-parallel field. The flow velocity perpendicular to the magnetic field (v_N) was nearly constant (except for a small shift of 5 km s^{-1}) across the bifurcated current sheet. The 5 km s^{-1} shift in v_N corresponds to a normal reconnection inflow $v_{N,\text{rec}}$ of 2.5 km s^{-1} (or a dimensionless reconnection rate, $v_{N,\text{rec}}/v_{\text{Alfven}}$ of 3.3%). The flow predictions in **c** and **d** are based on the local magnetic field measurements and the reference velocity and magnetic field: $v_{\text{predicted}} = v_{\text{reference}} \pm (1 - \alpha_{\text{reference}})^{1/2} (\mu_0 \rho_{\text{reference}})^{-1/2} [B_{\text{reference}}/\rho - B_{\text{reference}}]$ (refs 25, 26). The positive (negative) sign is chosen for the leading (trailing) edge of the bifurcated current sheet. $\alpha = (p_{\parallel} - p_{\perp})\mu_0/B^2$ is the pressure anisotropy factor, ρ is the plasma mass density. The left (right) dashed line in **c** and **d** denotes the reference times for the prediction of reconnection flow acceleration at the leading (trailing) edge of the exhaust. The leading and trailing edge predictions merge at 03:59 UT. The agreement between the predicted and observed flow is excellent in both the magnitude (**c**) and the components of the velocity (**d**). This level of agreement is similar to other reconnection exhaust events in the solar wind¹⁷.

more spacecraft most probably would not have encountered accelerated flow. Another fact that is consistent with a coherent and extended X-line is that the reconnection jets detected by all three spacecraft were directed in the same direction, implying that the X-line was north of all spacecraft. Patchy and random reconnection could result in different spacecraft detecting jets directed in different directions.

In addition to finding an extended X-line, the fact that the three spacecraft detected the reconnection exhaust over a period of 2.5 h implies that reconnection must have been quasi-steady over at least that time span. This finding is similar to reports of quasi-steady reconnection at the Earth's magnetopause^{22–24}. An important difference is that while reconnection is strongly driven at the magnetopause (by the solar wind impinging on the Earth's magnetosphere), reconnection in the present case appears to have been largely

undriven. There was a discontinuity in the flow speed across the current sheet of 27 km s^{-1} ; however, Fig. 3d shows that much of the flow speed discontinuity was due to a 22 km s^{-1} shear in the flow component tangential to the current sheet which does not compress the current sheet. In the normal direction the velocity across the current sheet was nearly constant except for a small 5 km s^{-1} shift. The velocity shift was consistent with a normal inflow, in the frame of the current sheet, of $v_{N,\text{rec}} = 2.5 \text{ km s}^{-1}$ associated with reconnection (at the position of Wind). The fact that reconnection can be quasi-steady in the undriven regime is surprising, and has not been previously reported to the best of our knowledge.

Finally, with a 12-nT magnetic field convecting into the reconnection region at 2.5 km s^{-1} (for a dimensionless reconnection rate, $v_{N,\text{rec}}/v_{\text{Alfven}}$ of 3.3%, where $v_{\text{Alfven}} = B/(\mu_0 \rho)^{1/2}$ is the Alfven speed), the reconnection electric field was 0.03 mV m^{-1} . Along an X-line of at least $390 R_E$, the minimum reconnection potential was thus 75 kV.

Although we have shown detailed observations from a single event, our conclusions in terms of extended X-lines and steady reconnection are general. We have identified 27 additional events when both ACE and Wind were in the solar wind and detected essentially the same reconnection signatures, irrespective of how far apart (in space and time) the two spacecraft were. Common among all 28 events is the fact that the plasma β (the ratio of plasma to magnetic pressure) in the ambient solar wind (outside the exhausts) is less than unity¹⁷ ($\langle \beta \rangle_{28 \text{ events}} = 0.4 \pm 0.2$), a condition that has been suggested to be necessary for the occurrence of reconnection²⁵. In four of these cases, we have evidence for an X-line extending more than $100 R_E$. We are aware of no counter-examples where one spacecraft detected the reconnection signature and the other did not. The large number of dual-spacecraft detections of reconnection flow with no counter-examples strongly indicate that reconnection in the solar wind, and probably in other astrophysical domains as well, can operate in a large-scale (much larger than the ion inertial scale) and quasi-steady mode, leading to the release of large amount of magnetic energy.

Our finding also raises an interesting question: how does the reconnection X-line become so extended? We suspect that in the case of the solar wind, reconnection starts in a limited region in the solar wind current sheet closer to the Sun and spreads with time from its initiation region. By the time the current sheet reaches 1 AU, the X-line has reached hundreds of R_E or more. The true size of the solar wind X-line can be investigated by the upcoming NASA/STEREO mission, which will provide large spacecraft separations that exceed 1 AU in the GSE-y direction.

Received 30 August; accepted 31 October 2005.

1. Pare, V. K. in *Magnetic Reconnection in Space and Laboratory Plasmas* (ed. Hones, E. W.) 341–346 (Geophysics Monograph 30, American Geophysical Union, Washington DC, 1984).
2. Cowley, S. W. H. in *Magnetic Reconnection in Space and Laboratory Plasmas* (ed. Hones, E. W.) 375–378 (Geophysics Monograph 30, American Geophysical Union, Washington DC, 1984).
3. Priest, E. R. in *Magnetic Reconnection in Space and Laboratory Plasmas* (ed. Hones, E. W.) 63–78 (Geophysics Monograph 30, American Geophysical Union, Washington DC, 1984).
4. Duncan, R. & Thompson, C. Formation of very strongly magnetized neutron stars: implications for gamma-ray bursts. *Astrophys. J.* **392**, L9–L13 (1992).
5. Hurley, K. et al. An exceptionally bright flare from SGR 1806–20 and the origins of short-duration γ -ray bursts. *Nature* **434**, 1098–1103 (2005).
6. Kronberg, P. P. Intergalactic magnetic fields. *Phys. Today* **55**, 40–46 (2002).
7. Nishida, A. Can random reconnection on the magnetopause produce the low latitude boundary layer? *Geophys. Res. Lett.* **16**, 227–230 (1989).
8. Russell, C. T. & Elphic, R. C. Initial ISEE magnetometer results: Magnetopause observations. *Space Sci. Rev.* **22**, 681–715 (1978).
9. Paschmann, G. et al. Plasma and magnetic characteristics of magnetic flux transfer events. *J. Geophys. Res.* **87**, 2159–2168 (1982).
10. Owen, C. J. et al. Cluster PEACE observations of electrons during magnetospheric flux transfer events. *Ann. Geophys.* **19**, 1509–1522 (2001).
11. Sonnerup, B., Hasegawa, H. & Paschmann, G. Anatomy of a flux transfer event seen by Cluster. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020134 (2004).

12. Nakamura, R. *et al.* Spatial scale of high-speed flows in the plasma sheet observed by Cluster. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL019558 (2004).
13. Louarn, P. *et al.* Cluster observations of complex 3D magnetic structures at the magnetopause. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020625 (2004).
14. Phan, T. D. *et al.* Extended magnetic reconnection at the Earth's magnetopause from detection of bi-directional jets. *Nature* **404**, 848–850 (2000).
15. Fuselier, S. A. *et al.* Cusp aurora dependence on IMF B_z . *J. Geophys. Res.* **107**, 1111, doi:10.1029/2001JA900165 (2002).
16. Pinnock, M. *et al.* The location and rate of dayside reconnection during an interval of southward interplanetary magnetic field. *Ann. Geophys.* **21**, 1467–1482 (2003).
17. Gosling, J. T. *et al.* Direct evidence for magnetic reconnection in the solar wind near 1AU. *J. Geophys. Res.* **100**, doi:10.1029/2004JA010809 (2005).
18. Gosling, J. T. *et al.* Magnetic disconnection from the Sun: Observations of a reconnection exhaust in the solar wind at the heliospheric current sheet. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL022406 (2005).
19. Petscheck, H. E. in *AAS-NASA Symp. on the Physics of Solar Flares* (28–30 October 1963, Goddard Space Flight Centre, Greenbelt, Maryland) (ed. Hess, W. N.) 425–437 (NASA Spec. Publ. SP-50, NASA Science and Technical Information Division, Washington DC, 1964).
20. Sonnerup, B. U. Ö. & Cahill, L. J. Jr Magnetopause structure and attitude from Explorer 12 observations. *J. Geophys. Res.* **72**, 171–183 (1967).
21. Sonnerup, B. U. Ö. Magnetopause reconnection rate. *J. Geophys. Res.* **79**, 1546–1549 (1974).
22. Gosling, J. T. *et al.* Evidence for quasi-stationary reconnection at the dayside magnetopause. *J. Geophys. Res.* **87**, 2147–2158 (1982).
23. Frey, H. *et al.* Continuous magnetic reconnection at Earth's magnetopause. *Nature* **426**, 533–537 (2003).
24. Phan, T. D. *et al.* Cluster observations of continuous reconnection at the magnetopause under steady interplanetary magnetic field conditions. *Ann. Geophys.* **22**, 2355–2367 (2004).
25. Paschmann, G. *et al.* The magnetopause for large magnetic shear: AMPTE/IRM observations. *J. Geophys. Res.* **91**, 11099–11115 (1986).
26. Hudson, P. D. Discontinuities in an anisotropic plasma and their identification in the solar wind. *Planet. Space Sci.* **18**, 1611–1622 (1970).

Acknowledgements Wind, ACE and Cluster research in the US is supported by NASA. Cluster research in France and the UK is supported by CNES and PPARC.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.D.P. (phan@ssl.berkeley.edu).

A semiconductor source of triggered entangled photon pairs

R. M. Stevenson¹, R. J. Young^{1,2}, P. Atkinson², K. Cooper², D. A. Ritchie² & A. J. Shields¹

Entangled photon pairs are an important resource in quantum optics¹, and are essential for quantum information² applications such as quantum key distribution^{3,4} and controlled quantum logic operations⁵. The radiative decay of biexcitons—that is, states consisting of two bound electron–hole pairs—in a quantum dot has been proposed as a source of triggered polarization-entangled photon pairs⁶. To date, however, experiments have indicated that a splitting of the intermediate exciton energy yields only classically correlated emission^{7–9}. Here we demonstrate triggered photon pair emission from single quantum dots suggestive of polarization entanglement. We achieve this by tuning the splitting to zero, through either application of an in-plane magnetic field or careful control of growth conditions. Entangled photon pairs generated ‘on demand’ have significant fundamental advantages over other schemes^{10–13}, which can suffer from multiple pair emission, or require post-selection techniques or the use of photon-number discriminating detectors. Furthermore, control over the pair generation time is essential for scaling many quantum information schemes beyond a few gates. Our results suggest that a triggered entangled photon pair source could be implemented by a simple semiconductor light-emitting diode¹⁴.

The most widely used methods for generating entangled photon pairs are nonlinear optical processes, such as parametric down conversion^{10,12}, which produce a probabilistic number of pairs per excitation cycle. Existing demonstrations of entangled photons in semiconductor systems also do not produce individual entangled photon pairs on demand. For the nonlinear process involved in bulk CuCl, the number of pairs emitted follows poissonian statistics¹⁵. For entangled photon pairs created by the probabilistic interference of indistinguishable photons from a single quantum dot¹⁶, post selection is required to reject the majority of photons which are not entangled¹³, and even for the idealized case only 50% of the photons are entangled. Thus the realization of a quantum dot source that emits no more than one entangled photon pair per excitation cycle is fundamentally different to the demonstrations described above. It is perhaps more closely related to generation of entangled photons in single atoms¹⁷, of which a quantum dot may be considered the semiconductor analogue. Another appealing feature is that after the first photon is emitted, the proposed single quantum dot system resides in an entangled photon–exciton state, opening up the possibility of implementing quantum logic operations in the solid state as well as photonic domains.

The radiative decay of the biexciton state (XX) in a quantum dot emits a pair of photons, with polarization determined by the spin of the intermediate exciton state (X). In an ideal quantum dot with degenerate X states, the polarization of the XX photon is predicted to be entangled with that of the X photon, forming the state $(|H_{XX}H_X\rangle + |V_{XX}V_X\rangle)/\sqrt{2}$, where H and V denote the polarization of the XX and X photons⁶.

In real quantum dots, the polarization of a photon can also be determined by its energy, due to splitting of the intermediate exciton state, shown schematically in Fig. 1. The splitting exists because of in-plane asymmetries of structural properties of the quantum dot, such as elongation and strain^{18,19}, and provides ‘which path’ information, preventing polarization entanglement of the emission. The key to the generation of entangled photon pairs in a quantum dot is therefore the reduction of the exciton polarization splitting to zero. Here we describe how we were able to carefully select unsplit quantum dots, or alternatively apply an in-plane magnetic field to tune the splitting to zero.

By characterizing the exciton polarization splitting of a large number of InAs/GaAs quantum dots from samples grown under different conditions, we found the splitting to be least for relatively

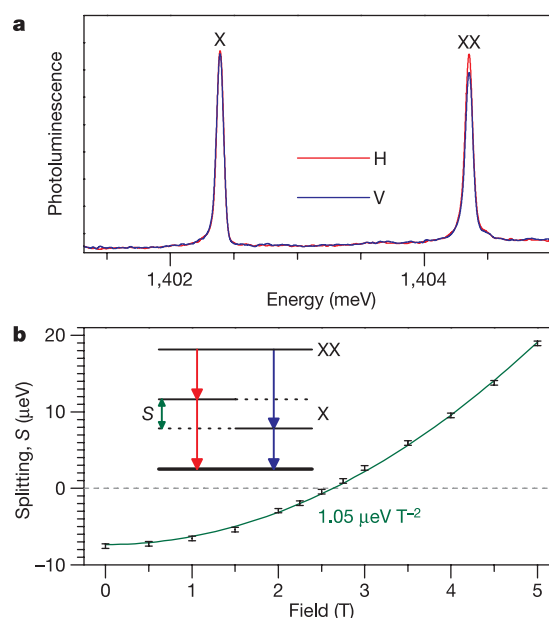


Figure 1 | Polarized photoluminescence spectra from single quantum dots.

a, Vertically (blue) and horizontally (red) polarized photoluminescence for a single quantum dot with small polarization splitting. The features correspond to emission by the exciton (X) and biexciton (XX) state.

b, Polarization splitting, S , as a function of in-plane magnetic field for a single dot with ‘inverted’ S at 0 T. The green line shows a quadratic fit to the data with a coefficient of $1.05 \mu\text{eV T}^{-2}$. Inset shows the level diagram of the radiative decay of the biexciton state. The competing two photon decay paths are distinguished only by the polarization of the photons, indicated by the arrow colour, and the splitting, S , of the intermediate exciton level. Error bars span two standard deviations from the fitted line.

¹Toshiba Research Europe Limited, 260 Cambridge Science Park, Cambridge CB4 0WE, UK. ²Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK.

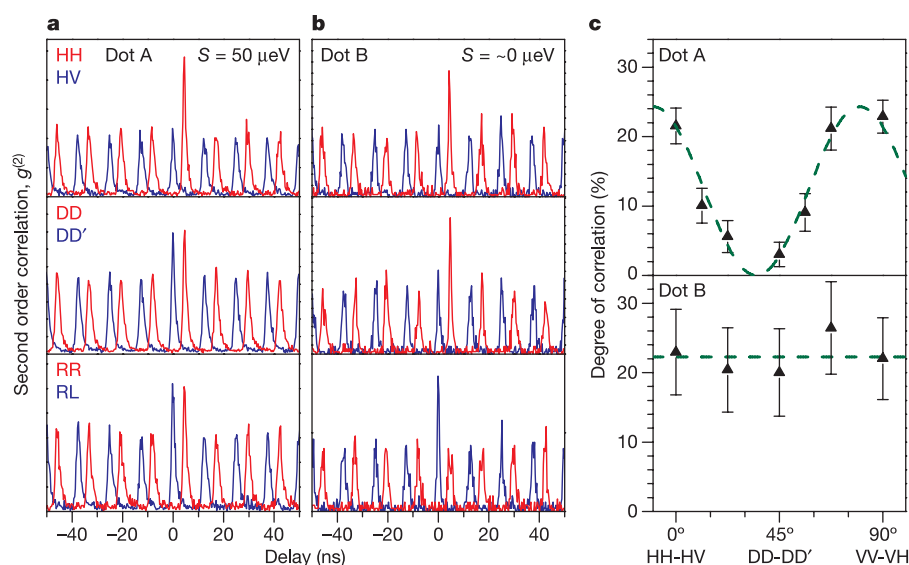


Figure 2 | Second order cross correlation of biexciton with exciton photons from conventional and degenerate single quantum dots. **a, b,** Cross correlation for a reference dot A with 50 μeV polarization splitting (**a**), and a degenerate dot B (**b**). Correlations measured for photons of the same polarization are shown in red, and for orthogonal polarization in blue. The red histograms are time shifted to allow easier comparison to the blue. The

top, middle and bottom panels represent correlations measured in the rectilinear, diagonal and circular bases, respectively. **c,** The degree of linear correlation is plotted as a function of the basis angle. Error bars span two standard deviations. H, horizontal; V, vertical; D, diagonal; D', orthodiagonal; R, right; L, left.

small dots emitting at $\sim 1.4\text{ eV}$, in which the electron and hole wavefunctions are most symmetric²⁰. Figure 1a plots polarized photoluminescence spectra at $\sim 10\text{ K}$, recorded for a typical dot showing near-zero splitting. The biexciton photon emission energy is 1.96 meV higher than the exciton photon emission energy, which is quite typical for this kind of small quantum dot^{21,22}. The energy difference, due to the Coulomb interaction, allows simple wavelength selective separation of the two photons in the pair, which is not possible in most schemes for entangled photon generation that require the photons to have the same energy. No polarization splitting can be resolved by eye between the horizontally and vertically polarized components of the emission. The fitted central wavelengths of all lines were used to determine the average polarization splitting for the exciton and biexciton. The resulting measured exciton splitting was found to be $1.1 \pm 0.5\text{ }\mu\text{eV}$, within the projected homogeneous linewidth of $\sim 1.5\text{ }\mu\text{eV}$.

The samples show considerable difference in the magnitude and sign of splitting from dot to dot, owing to variations in dot size, shape and composition. However, we have found that the polarization splitting of many dots can be tuned to zero by applying an in-plane magnetic field. These dots are identified by the negative sign of their splitting, using the sign convention that the splitting equals the energy of the horizontally polarized exciton minus that of the vertically polarized one. The tuning of the polarization splitting of an example dot by magnetic field is shown in Fig. 1b. The splitting varies approximately quadratically with magnetic field from $-7.6 \pm 0.5\text{ }\mu\text{eV}$ to $+19.3 \pm 0.5\text{ }\mu\text{eV}$. Most importantly, the splitting can be tuned to approximately zero for an applied field of $2.5 \pm 0.1\text{ T}$. As we shall show later, this provides a way to turn on entangled photon generation by the quantum dot.

Polarization dependent two photon correlations were measured as detailed in the Methods section. Figure 2a shows the second order cross correlation between the XX and X photon emitted by a reference dot A, for which the exciton splitting was $49.9 \pm 0.5\text{ }\mu\text{eV}$. The polarization correlated traces shown in red are artificially time shifted relative to the polarization anti-correlated traces shown in blue, to aid comparison. The average total number of counts in the peaks at zero delay is larger than the average number of counts in the other peaks. This is due to the greater probability of detecting an

X photon in the same period that an XX photon was detected. The degree of enhancement is dependent on the excitation efficiency, which is the same for each pair of traces recorded simultaneously, allowing direct comparison. The three panels represent correlations measured in the rectilinear, diagonal and circular bases, and were measured using a half wave plate set at 0° and 22.5° , and a quarter wave plate set at 45° , placed directly after the collection lens. The top panel shows a much higher probability of generating a pair of photons with the same rectilinear polarization. Detection of oppositely polarized photon pairs is not fully suppressed owing to contribution from background light and exciton dephasing, discussed later in detail. However, no polarization correlation at all is observed in the diagonal or circular bases, demonstrated by the middle and bottom panels. The photon pairs emitted by the reference dot are therefore only classically polarization correlated as observed previously⁷⁻⁹, and not entangled.

The correlation experiment was repeated for dot B, which has approximately zero splitting. The results are shown in Fig. 2b. A similar degree of rectilinear polarization correlation is observed as for the reference dot A, and indeed all dots measured. However, strong diagonal polarization correlation is additionally observed, and remarkably also strong circular polarization anti-correlation. This is consistent with expected results for entangled photon emission from a single quantum dot.

A well known property of entangled photon pairs is that the degree of correlation does not depend on the absolute orientation of the photons, but only the difference in the angle between them¹². The degree of linear correlation was measured as a function of the linear polarization basis angle by rotating the half wave plate, which has no effect on the difference in polarization detection angles, which was effectively parallel. The results are shown in Fig. 2c. For the reference dot A, the correlation fits well to sinusoidal behaviour between zero and a maximum of 0.243 ± 0.012 , as expected for classically linearly polarization correlated photon pairs. In stark contrast, the degree of correlation for the degenerate dot B is independent of the orientation of the measurement basis within experimental error, as expected for polarization entangled photon pairs. The average degree of correlation is found to be 0.222 ± 0.028 , similar to the maximum of the reference dot. Again this is what is expected for entangled photon

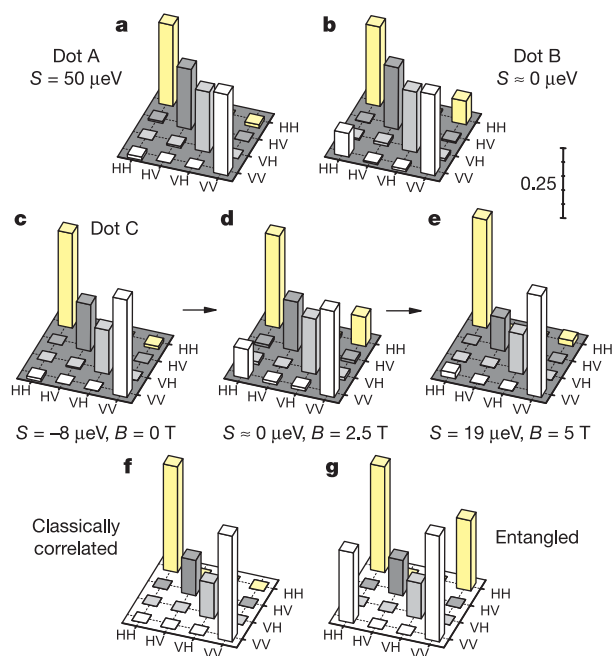


Figure 3 | Density matrices for the biexciton-exciton two-photon cascade from conventional and degenerate quantum dots. **a–e**, Real parts of measured density matrices corresponding to reference dot A with polarization splitting, $S = 50 \mu\text{eV}$ (**a**), dot B with $S \approx 0 \mu\text{eV}$ at 0 T (**b**), and dot C, with S tuned by the magnetic field to be $-8 \mu\text{eV}$ (**c**), $0 \mu\text{eV}$ (**d**) and $19 \mu\text{eV}$ (**e**). The imaginary components are not shown, and were zero within experimental error. Density matrices **b** and **d** feature strong outer off diagonal elements associated with entangled photon pair states, which are not present in the reference case (**a**). **f**, **g**, Density matrices representing the predicted state for ideal classically correlated (**f**) and entangled (**g**) photon pairs, including 50% contribution from uncorrelated background light.

pairs, in contrast to the case of classically correlated photon pairs emitting into random bases, for which the average degree of correlation should be less by a factor of 2.

To fully measure the two photon polarization state, a quantum state tomography scheme was used^{23,24}. The procedure, detailed in the Methods section, constructs the two photon polarization density matrix from a linear combination of cross correlation measurements using 16 different polarization combinations.

The real component of the two photon polarization density matrix for the reference dot A is shown in Fig. 3a. The stronger elements all lie on the diagonal, with the strongest outer elements indicating polarization correlated emission. The inner diagonal components are due to uncorrelated photon pair emission, from background counts and dephasing of the exciton state. The form of this density matrix is consistent with imperfect polarization correlated photon pair emission seen previously^{7–9}, and illustrated by the example density matrix of Fig. 3f. The density matrix for the degenerate dot shown in Fig. 3b has similar diagonal elements, but now shows significant outer, off diagonal elements. This is a feature associated with polarization entangled photon pairs, illustrated by the predicted density matrix of Fig. 3g.

A similar density matrix is obtained for dot C, tuned to zero splitting by magnetic field, as shown in Fig. 3d. This again suggests that the photon pair emission has entangled character. When the field is increased to 5 T, the splitting increases to $19 \mu\text{eV}$, and the corresponding density matrix measured is shown in Fig. 3e. As expected, the off diagonal elements are suppressed, and the dot reverts to emitting polarization correlated photon pairs. A similar result is found if the field is reduced to 0 T, where the splitting is $-8 \mu\text{eV}$ as shown in Fig. 3c. The imaginary components of the density matrices

were all found to be zero with experimental error, in agreement with predictions.

The measurements presented above clearly suggest that dots with small exciton splitting emit entangled photons. We now discuss the factors limiting the degree of entanglement. In spectroscopy, our measurements show that the background due to dark counts and emission from layers other than the dot contributes on average 49% of the coincidence counts; this is unusually large owing to the proximity of the dot to the wetting layer, which is necessary to select dots with zero splitting. If we correct our measurements by removing the projected number of background counts from the correlation data, the density matrices of the degenerate and magnetically tuned dots more closely resemble the ideal entangled case, and the largest eigenvalues are 0.48 ± 0.08 and 0.58 ± 0.04 , respectively. The latter, for which the splitting is minimal, violates the 0.5 limit for classical correlation in an unpolarized source²⁵. The remaining deviation from ideal behaviour is attributed to scattering between the two intermediate exciton spin states^{7,8}. From previous publications where strong background and entanglement were not present, we estimate an exciton scattering time similar to the $\sim 1 \text{ ns}$ radiative lifetime. This yields a maximum possible eigenvalue of 0.63, in rough agreement with these measurements.

This suggests that the degree of entanglement may be increased by resonant optical^{16,26} or electrical⁶ excitation in order to increase the scattering time, or by reducing the radiative lifetime through Purcell enhancement^{27,28}, or by using dots with larger oscillator strength such as those formed by interface fluctuations¹⁸. Such improvements could lead to the realization of a semiconductor source of triggered entangled photon pairs that would be robust and compact, and allow electrical injection of the carriers¹⁴.

METHODS

Sample fabrication and characterization. Samples containing a low density layer of InAs quantum dots (~ 1.6 monolayers thick) were grown by molecular beam epitaxy. A GaAs λ cavity containing the dot layer was surrounded by AlAs/GaAs distributed Bragg reflectors, with 14 (2) repeats in the bottom (top) mirror, to increase light collection efficiency. A metal shadow mask containing apertures of $\sim 2 \mu\text{m}$ diameter was fabricated to isolate the emission of individual dots. Samples with a range of InAs thicknesses differing by up to $\sim 2\%$ were characterized in a standard micro photoluminescence system operating at $\sim 10 \text{ K}$. Optical excitation was provided by $\sim 100\text{-ps}$ pulses from a 635-nm laser diode operating at 80 MHz. The emission lines are inhomogeneously broadened by charge fluctuations to $\sim 50 \mu\text{eV}$, a consequence of the non-resonant excitation scheme²⁹. Horizontally ($[110]$) and vertically ($[1\bar{1}0]$) polarized exciton and biexciton emission was fitted with lorentzian line shapes to locate the centre energy of each transition. The exciton level splitting can be determined both from the difference between the horizontally and vertically polarized exciton or biexciton photons. Taking the average of these two values removed systematic error associated with changing the polarization optics, and the splitting S was measured with an estimated precision of $\sim 0.5 \mu\text{eV}$.

Selection of suitable dots. By measuring the splitting of 200 quantum dots, a relationship of decreasing splitting with increasing emission energy was found²⁰. For dots emitting at $\sim 1.4 \text{ eV}$, the splitting was $\sim 0 \pm 10 \mu\text{eV}$. Thus quantum dots with splitting less than the homogeneous linewidth of $\sim 1.5 \mu\text{eV}$ were selected first by identifying dots emitting close to 1.4 eV , then measuring their splitting. For dots emitting $> 1.4 \text{ eV}$, the splitting was inverted, and the lowest energy exciton line is horizontally polarized. For these dots, the configuration of the exchange energies and g -factors allows reduction of the splitting with an applied in-plane magnetic field, driven by partial mixing of optically active and inactive exciton states³⁰. Thus dots suitable for tuning to zero splitting are conveniently identified by their emission energy. The proportion of dots that have, or can be tuned to, zero splitting is $\sim 30\%$, which could be improved by better growth control. The proportion of suitably isolated single dots could be improved by fabrication of smaller microstructures.

Photon pair counting. Quantum dots were optically excited, with the power adjusted to give optimum photon pair detection rate to background ratio. At this power, the biexciton intensity is around half that of the exciton. A 50/50 beam splitter divided the emission into two spectrometers, set to transmit at the XX and X photon energies respectively, with $\sim 0.5 \text{ meV}$ bandwidth. Polarizing beam splitters were placed after the spectrometers, and three single photon detectors

were used to measure the vertically polarized XX photons, and horizontally and vertically polarized X photons. The time between detection of XX photons and X photons was measured by a time interval analyser.

Photons can be counted over a number of hours by compensating for fluctuations in excitation and detection efficiency over time. This is achieved by determining the degree of correlation from the ratio of the two correlations measured simultaneously, each normalized by the number of pairs detected in different laser cycles.

The approach is valid for unpolarized sources, and unpolarizing transmission of the light collection system up to the wave plates. Our system satisfies these requirements, as the emission was unpolarized within error, and the transmission was only weakly polarizing. For the results of Fig. 2, a single half or quarter wave plate was inserted directly after the collection lens to select the measurement basis for the exciton and biexciton photons simultaneously, and the transmission was zero within experimental error. For the measurements of Fig. 3, a quarter wave and half wave plate was used before each spectrometer to select the polarization detection basis for the exciton and biexciton independently, and the transmission was only ~10% polarizing. The total number of coincident pairs detected over the course of an experiment is typically up to 1,000, which dictates the measurement errors.

Quantum tomography analysis. The probability that a photon pair is detected with a selected polarization combination was determined experimentally from pairs of two photon correlations measured simultaneously. 16 such measurements were used to construct each two photon density matrix, using the polarization combinations and methods of ref. 24. The density matrix fully describes the two-photon quantum state, and thus can be used to test for entanglement. The test we chose is that of the largest eigenvalue, which is a general test that makes no assumption about the nature of the entangled state, in contrast to a Bell inequality, for which violation is maximal when the entangled state corresponds to that for which the polarization measurement bases were chosen. For an unpolarized classical source, the probability that a photon pair exists in any given polarization state cannot exceed 0.5, therefore an eigenvalue >0.5 signifies the presence of entanglement.

Received 28 July; accepted 16 November 2005.

- Walls, D. F. & Milburn, G. J. *Quantum Optics* (Springer, Berlin, 1994).
- Bouwmeester, D., Ekert, A. K. & Zeilinger, A. *The Physics of Quantum Information* (Springer, Berlin, 2000).
- Ekert, A. K. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991).
- Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195 (2002).
- Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
- Benson, O., Santori, C., Pelton, M. & Yamamoto, Y. Regulated and entangled photons from a single quantum dot. *Phys. Rev. Lett.* **84**, 2513–2516 (2000).
- Stevenson, R. M. *et al.* Quantum dots as a photon source for passive quantum key encoding. *Phys. Rev. B* **66**, 081302 (2002).
- Santori, C., Fattal, D., Pelton, M., Solomon, G. S. & Yamamoto, Y. Polarization-correlated photon pairs from a single quantum dot. *Phys. Rev. B* **66**, 045308 (2002).
- Ulrich, S. M., Strauf, S., Michler, P., Bacher, G. & Forchel, A. Triggered polarization-correlated photon pairs from a single CdSe quantum dot. *Appl. Phys. Lett.* **83**, 1848–1850 (2003).
- Shih, Y. H. & Alley, C. O. New type of Einstein-Podolsky-Rosen-Bohm experiment using pairs of light quanta produced by optical parametric down conversion. *Phys. Rev. Lett.* **61**, 2921–2924 (1988).
- Ou, Z. Y. & Mandel, L. Violation of Bell's inequality and classical probability in a two-photon correlation experiment. *Phys. Rev. Lett.* **61**, 50–53 (1988).
- Kiess, T. E., Shih, Y. H., Sergienko, A. V. & Alley, C. O. Einstein-Podolsky-Rosen-Bohm experiment using pairs of light quanta produced by type-II parametric down-conversion. *Phys. Rev. Lett.* **71**, 3893–3897 (1993).
- Fattal, D. *et al.* Entanglement formation and violation of Bell's inequality with a semiconductor single photon source. *Phys. Rev. Lett.* **92**, 037903 (2004).
- Yuan, Z. *et al.* Electrically driven single-photon source. *Science* **295**, 102–105 (2002).
- Edamatsu, K., Oohata, G., Shimizu, R. & Itoh, T. Generation of ultraviolet entangled photons in a semiconductor. *Nature* **431**, 167–170 (2004).
- Santori, C., Fattal, D., Vučković, J., Solomon, G. S. & Yamamoto, Y. Indistinguishable photons from a single-photon device. *Nature* **419**, 594–597 (2002).
- Aspect, A., Grangier, P. & Roger, G. Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A new violation of Bell's inequalities. *Phys. Rev. Lett.* **49**, 91–94 (1982).
- Gammon, D., Snow, E. S., Shanabrook, B. V., Katzer, D. S. & Park, D. Fine structure splitting in the optical spectra of single GaAs quantum dots. *Phys. Rev. Lett.* **76**, 3005–3008 (1996).
- Bimberg, D., Grundmann, M. & Ledentsov, N. N. *Quantum Dot Heterostructures* (Wiley, Chichester, 1999).
- Young, R. J. *et al.* Inversion of exciton level splitting in quantum dots. *Phys. Rev. B* **72**, 113305 (2005).
- Thompson, R. M. *et al.* Single-photon emission from exciton complexes in individual quantum dots. *Phys. Rev. B* **64**, 201302 (2001).
- Rodt, S. *et al.* Repulsive exciton-exciton interaction in quantum dots. *Phys. Rev. B* **68**, 035331 (2003).
- White, A. G., James, D. F. V., Eberhard, P. H. & Kwiat, P. G. Nonmaximally entangled states: Production, characterization, and utilization. *Phys. Rev. Lett.* **83**, 3103–3107 (1999).
- James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001).
- Kwiat, P. G., Barraza-Lopez, S., Stefanov, A. & Gisin, N. Experimental entanglement distillation and 'hidden' non-locality. *Nature* **409**, 1014–1017 (2001).
- Santori, C., Pelton, M., Solomon, G., Dale, Y. & Yamamoto, Y. Triggered single photons from a quantum dot. *Phys. Rev. Lett.* **86**, 1502–1505 (2001).
- Pelton, M. *et al.* Efficient source of single photons: A single quantum dot in a micropost microcavity. *Phys. Rev. Lett.* **89**, 233602 (2002).
- Bennett, A. J., Unitt, D. C., Atkinson, P., Ritchie, D. & Shields, A. J. High performance single photon sources from photolithographically defined pillar microcavities. *Opt. Express* **13**, 50–55 (2005).
- Bayer, M. & Forchel, A. Temperature dependence of the exciton homogeneous linewidth in $\text{In}_{0.60}\text{Ga}_{0.40}\text{As}$ /GaAs self-assembled quantum dots. *Phys. Rev. B* **65**, 041308 (2002).
- Stevenson, R. M. *et al.* Magnetic-field-induced reduction of the exciton polarisation splitting in InAs quantum dots. *Phys. Rev. B* (in the press).

Acknowledgements We acknowledge continued support from M. Pepper. This work was partially funded by the EU projects RAMBOQ, QAP and SANDiE, and by the EPSRC through the IRC for Quantum Information Processing.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.M.S. (mark.stevenson@crl.toshiba.co.uk).

Ultrafast superheating and melting of bulk ice

H. Iglev¹, M. Schmeisser¹, K. Simeonidis¹, A. Thaller¹ & A. Laubereau¹

The superheating of a solid to a temperature beyond its melting point, without the solid actually melting, is a well-known phenomenon. It occurs with many substances^{1–5}, particularly those that can readily be produced as high-quality crystals. In principle, ice should also be amenable to superheating. But the complex three-dimensional network of hydrogen bonds that holds water molecules together and gives rise to unusual solid and liquid properties^{6–11} strongly affects the melting behaviour of ice^{12–14}; in particular, ice usually contains many defects owing to the directionality of its hydrogen bonds. However, simulations are readily able to ‘create’ defect-free ice that can be superheated^{15,16}. Here we show that by exciting the OH stretching mode of water, it is possible to superheat ice. When using an ice sample at an initial temperature of 270 K, we observe an average temperature rise of 20 ± 2 K that persists over the monitored time interval of 250 ps without melting.

We have performed ultrafast temperature jump measurements in ice (for details on the experimental set-up and the sample preparation, see the Methods section) using the OH- and OD-stretching vibrations for rapid heating of the sample^{17,18}. The same vibrational modes are known to be sensitive probes for H-bonding^{19,20} and represent suitable spectral tools to distinguish local ice or water structures with a time resolution of a few picoseconds^{20,21}. To illustrate this point, the steady-state infrared absorption spectrum of HDO:D₂O is shown in Fig. 1a for various temperature values. In the range 2,000–4,000 cm^{−1} the well-known OD- and OH-bands (left and right) occur with significant changes of position (blueshift) and shape with temperature. Figure 1b presents the same data as thermal differential spectra for $\Delta T = 20$ K. The spectrum corresponding to melting of the ice sample reduced by a factor of five is also shown in the Fig. 1b. The potential of the molecular vibrations as local probes of temperature and melting is readily seen.

Examples for the time-resolved differential spectra measured in a 2.5- μ m-thick HDO:D₂O (15 M) ice specimen at 200 K are presented in Fig. 2. It shows ultrafast heating by two different processes: via infrared absorption of the OH- or the OD-stretching vibration. The transient spectra taken 50 ps after excitation are shown in Fig. 2a. The absorption changes are plotted for pump pulses of 3.0 μ J at 3,290 cm^{−1} (OH-pumping, blue points) and 2.7 μ J at 2,435 cm^{−1} (OD-pumping, red points). The adjustment of the pump-pulse energies ensures that approximately equal amounts of energy are deposited in the sample for both frequency settings. Both excitation schemes induce almost the same spectral changes in the HDO ice, indicating that the OH- and OD-oscillators are already at $t_D = 50$ ps in a local equilibrium. The similarity of the time-resolved data in Fig. 2a and the thermal differential spectra presented in Fig. 1b confirms induced heating of the sample. A comparison of the experimental data against differential spectra computed for various isochoric temperature jumps and the associated simultaneous pressure increase suggests an average temperature rise of 20 ± 2 K in the ice sample (Fig. 2b, see also the Methods section for

further details). The isochoric pressure increase is estimated to be ~ 26 MPa.

The time evolution of the spectral changes induced by OH-pumping is depicted as a contour plot in Fig. 2c, showing that the frequency shift of the hydroxilic stretching modes—which appears as induced absorption in the high-frequency wings, indicated in red—has terminated after 20 ps. The temporal evolution of the spectral shift for both OH- and OD-pumping is documented by more quantitative data in Fig. 2d and e, for fixed probing frequencies of 2,435 cm^{−1} (OD probing, Fig. 2d) and 3,280 cm^{−1} (OH probing, Fig. 2e). At short delay times, fast signal changes occur at rates comparable to our time resolution under both excitation conditions, at both probing frequencies. (The signal overshoot for OH-pumping, shown by blue points in Fig. 2e at around $t_D = 0$ ps possibly involves a coherent pump-probe artefact in addition to depletion of the vibrational ground state.) For delay times longer than 20 ps, amplitude levels are constant. We can fit the data using a simple relaxation model with two assumed exponential time constants, and obtain a short relaxation time of $\tau_1 \leq 0.5$ ps consistent with the reported OH-lifetime of HDO ice^{17,18} and a longer relaxation time of

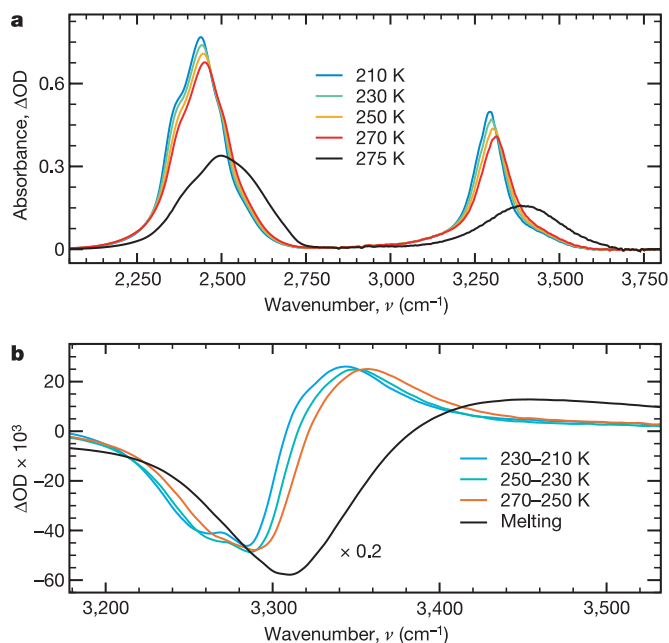


Figure 1 | Conventional infrared absorption spectra of HDO:D₂O (15 M). **a**, Absorption spectra of hexagonal ice in the OH- and OD-stretching region at various temperatures from 210 K (blue) to 270 K (red). The spectrum of the molten sample at 275 K (black line) is also shown. **b**, Same data as in **a** plotted as thermal differential spectra with $\Delta T = 20$ K. For a better view, the spectrum corresponding to melting of the ice sample is scaled by a factor of 0.2 (275–270 K, black line).

¹Physik-Department, Technische Universität München, James-Frank-Strasse, D-85748 Garching, Germany.

$\tau_2 = 8.7 \pm 0.5$ ps attributed to the thermalization of the hydrogen bonding network of the ice specimen, that is, energy redistribution among low-frequency intermolecular vibrations. The process involves an average temperature rise of 20 ± 2 K that persists beyond 250 ps.

Figure 3 presents data similar to those shown in Fig. 2, except that the ice sample has an initial temperature of 270 K and that the pump frequencies are tuned to match the temperature-shifted band maxima ($3,310\text{ cm}^{-1}$, blue points; $2,450\text{ cm}^{-1}$, red points). The time-resolved differential spectra for $t_D = 50$ ps in Fig. 3a show a striking similarity to the spectra in Fig. 2a, exhibiting a blueshift of the hydroxilic bands but no evidence for melting. Similarly, the time evolution of the probe signals (Fig. 3c–e) only suggests an increase in the temperature of the sample that does not cause melting. The energy transfer between OH- and OD-modes occurs slightly faster than at 200 K, with the new thermal equilibrium established within 15 ps. The behaviour is reflected by a shorter thermal relaxation time obtained with the two-stage relaxation model, $\tau_2 = 5.4 \pm 0.5$ ps (Fig. 3d and e; we can again place only an upper limit on the fast relaxation time, $\tau_1 \leq 0.5$ ps).

The similar behaviour observed at the two temperatures suggests that the optical excitation induces the same process in both ice samples: heating of the ice, rather than melting. To verify this point, we numerically generate a hypothetical thermal differential spectrum for HDO:D₂O ice at a temperature that is above the ice melting point of 274.8 K. For this, we extrapolate the dependence of the spectral position of the $\Delta T = 20$ K thermal difference spectrum on initial

sample temperature, noting that both amplitude and spectral shape depend only slightly on initial temperature (Fig. 1b). The results of the simulation for different temperature jumps (which account for the effects of isochoric pressure increases) are compared with the measured spectrum in Fig. 3b, indicating that optical excitation has heated the 270 K ice sample to an averaged bulk temperature of 290 ± 2 K.

According to simulations, the stability of the superheated state of metals is limited by defect concentrations of the order of several per cent²², whereas experimental studies observed melting already at a vacancy concentration one order of magnitude smaller²³. We estimate the chemical impurities of our sample to be of the order of 10^{-4} , representing a lower limit to the defect concentration. The strength of hydrogen bonds is notably smaller than that of covalent and metallic bonds, resulting in higher fluctuations of the hydrogen-bonded network in ice and thus enhanced formation and migration of structural defects²⁴; in the case of large topological defects with a lifetime of 0.5 ns, these are predicted to induce bulk melting close to the common melting temperature²⁵. In general, defect-free crystals can readily be superheated^{15,16}, but once a critical defect concentration is present, it will induce melting and thus interfere with superheating²⁵. We note that evidence for the melting of our samples within 250 ps after energy deposition is lacking. The topological defects considered in ref. 25 clearly do not play a significant role on this timescale.

When using larger temperature jumps, we do observe melting. Figure 4 shows time-resolved data for OH-pumping at $3,275\text{ cm}^{-1}$, obtained with a $1\text{-}\mu\text{m}$ -thick sample of HDO:H₂O (15 M) ice with an initial temperature of 265 K. The absorption changes at four spectral positions in the OH-stretching region measured at $t_D = 50$ ps are plotted versus density of deposited energy (Fig. 4a). As long as the deposited energy density does not

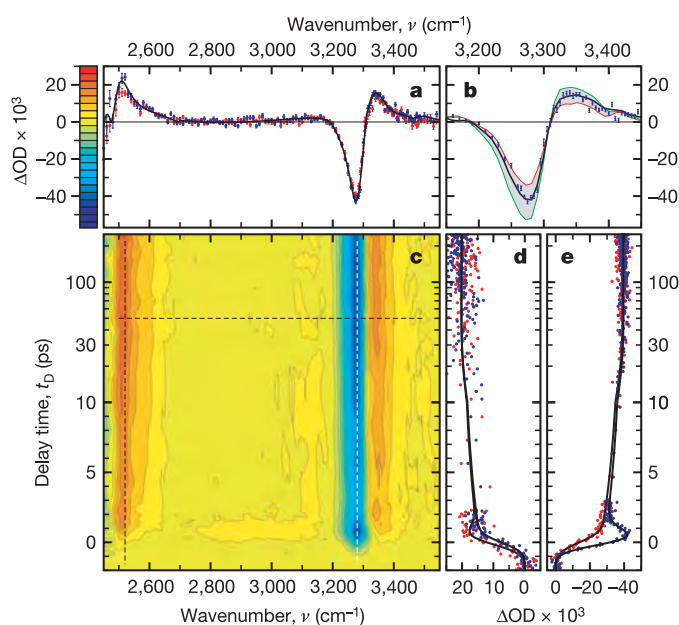


Figure 2 | Ultrafast heating of HDO:D₂O ice from 200 K to 220 K.

Subpicosecond excitation at $3,290\text{ cm}^{-1}$ (OH-pumping, blue points) and $2,435\text{ cm}^{-1}$ (OD-pumping, red points). Error bars are s.d. **a**, Transient differential absorption spectra ($\Delta\text{OD} \times 10^3$) taken 50 ps after excitation. The solid black line indicates the differential absorption spectrum calculated for a temperature increase of 20 K. **b**, Magnification of the OH-range from panel **a**, with calculated isochoric thermal differential spectra for $\Delta T = 15$ K (red), 20 K (black) and 25 K (green). **c**, Contour plot of the spectral changes induced by OH-pumping as a function of probe frequency and delay time. Induced absorption is indicated by red and bleaching by blue, with contour values given in the colour scale next to **a**. We note that the delay time is on a linear scale to 10 ps, and on a logarithmic scale for longer delays. **d**, **e**, Induced absorption ($\Delta\text{OD} \times 10^3$) at $2,520\text{ cm}^{-1}$ (**d**) and bleaching (negative $\Delta\text{OD} \times 10^3$) at $3,280\text{ cm}^{-1}$ (**e**) versus delay time. Solid lines are fits using a simple relaxation model with two assumed exponential time constants.

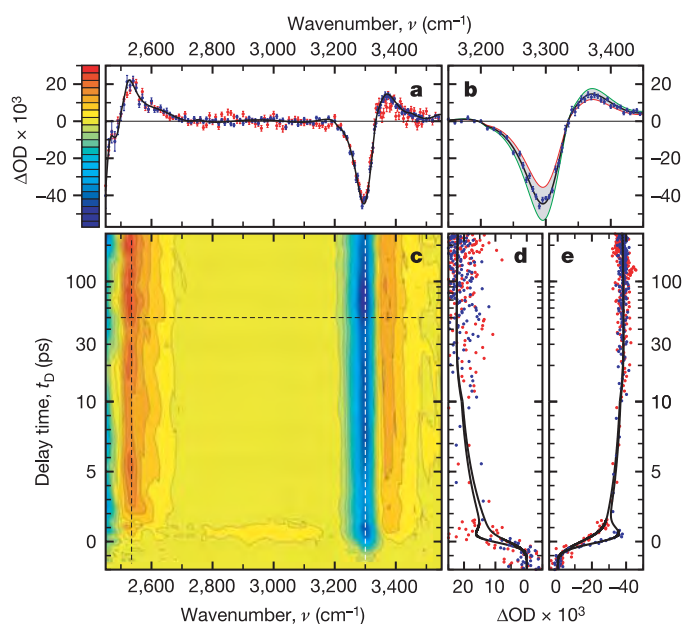


Figure 3 | Superheating of HDO:D₂O ice from 270 K to 290 K. Excitation at $3,310\text{ cm}^{-1}$ (blue points) and $2,450\text{ cm}^{-1}$ (red points). Error bars are s.d. **a**, Transient differential absorption spectra measured 50 ps after ultrafast heating. The black solid curve represents an extrapolated thermal differential absorption spectrum for a transient ice state at 290 K and pressure of 26 MPa. **b**, Magnification of the OH-range with extrapolated isochoric thermal differential absorption spectra for $\Delta T = 15$ K (red), 20 K (black) and 25 K (green). **c**, Contour plot of the spectral changes induced by OH-pumping, as a function of probe frequency and delay time. **d**, **e**, Induced absorption at $2,535\text{ cm}^{-1}$ (**d**) and bleaching at $3,300\text{ cm}^{-1}$ (**e**) versus delay time.

exceed $55 \pm 5 \text{ J cm}^{-3}$, the spectral amplitudes are proportional to the excitation level (and average temperature rise), as is expected for laser-induced heating. Similar to the data of Fig. 3 for the HDO:D₂O sample, the superheated ice state of HDO:H₂O is found to persist over the monitored time interval of 250 ps (data not shown).

When the deposited energy density exceeds 55 J cm^{-3} , all excitation curves exhibit a change in their slope (see Fig. 4a) that is accompanied by a significant broadening of the transient absorption spectra. The difference of two transient spectra measured above 55 J cm^{-3} agrees well with the thermal differential spectrum corresponding to the ice–liquid phase transition (data not shown), suggesting that the sample undergoes partial melting. The data in Fig. 4a suggest an average temperature value of $292 \pm 5 \text{ K}$ as the limit for superheating of this ice sample, a value in fair agreement with recent calculations by Luo *et al.*¹⁵

Data on the time evolution of the melting process in the HDO:H₂O sample after excitation with $95 \pm 5 \text{ J cm}^{-3}$ are presented in Fig. 4b. For short delay times, absorption first decreases rapidly (see inset of Fig. 4b) and then more slowly on the timescale of several picoseconds (τ_2). In contrast to the dynamics observed for superheating (below 55 J cm^{-3}), where the signal curve reaches a constant amplitude within 15 ps, the data in Fig. 4b display an additional

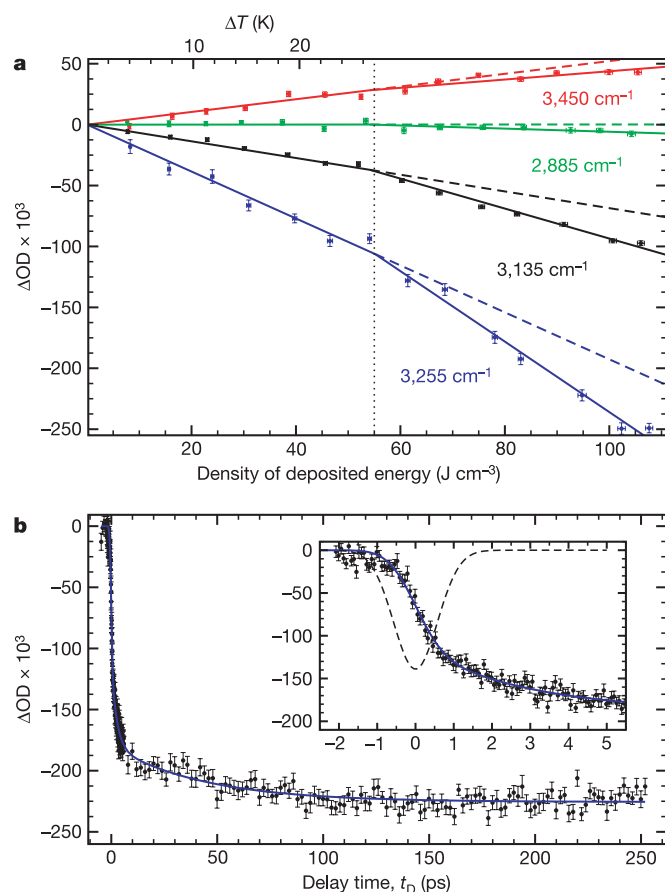


Figure 4 | Superheating and (partial) melting of ice. Time-resolved absorption changes ($\Delta OD \times 10^3$) measured in HDO:H₂O (15 M) ice at 265 K after OH-pumping at $3,275 \text{ cm}^{-1}$ (experimental points, error bars are s.d.). **a**, Absorption changes taken at $t_D = 50 \text{ ps}$ and various spectral positions versus average density of deposited energy in the probed volume; lines are guides to the eye. The upper abscissa scale denotes the average temperature jump ΔT . **b**, Temporal evolution of the induced bleaching in the OH-band at $3,255 \text{ cm}^{-1}$; the solid curve indicates the predictions of the extended relaxation model. The inset shows the short time dynamics ($\Delta OD \times 10^3$ versus t_D) together with the shape of the pump pulse (dashed line).

relaxation mode extending over 150 ps that appears to be connected to partial melting of the ice specimen. Incorporating a third exponential time constant in the relaxation model yields $\tau_3 = 45 \pm 5 \text{ ps}$ for the slowest process (see solid curves in Fig. 4b). The exact character of this time constant is not well understood at the present time. Compared to the lifetime of hydroxilic vibrations (below 1 ps) and the measured thermalization time of ice of several picoseconds, this novel relaxation feature persists over a notably longer time period; this is expected for a collective process such as melting, which involves many molecules. Deeper understanding of this intriguing feature will need to await more detailed studies of the mechanisms underlying the solid–liquid phase transition.

METHODS

Subpicosecond infrared spectrometer. Our experimental system was described recently²⁶ and is just briefly mentioned here. The infrared pulses are delivered by laser-pumped parametric oscillator-amplifier devices with durations of 0.7 ps (0.9 ps), spectral widths of 24 cm^{-1} (19 cm^{-1}) and typical energies of 10 nJ ($3.0 \mu\text{J}$). Computer-controlled tuning is provided in the range $1,700\text{--}3,700 \text{ cm}^{-1}$ ($2,300$ to $3,700 \text{ cm}^{-1}$). Numbers in brackets refer to the pump pulses. The pump beam diameter in the sample of approximately $150 \mu\text{m}$ is a factor of two larger than that of the probe, so that only the central part of the interaction volume with maximum excitation is monitored. The energy transmittances $T(v)_\parallel$, $T(v)_\perp$ of the probing pulse through the excited sample are measured for parallel ($\parallel$) and perpendicular ($\perp$) polarizations with respect to the linear polarization of the pump beam and compared with the probe transmittance $T(v)_0$ for blocked excitation. In this way the induced changes of optical density $\Delta OD_{\parallel,\perp}$ are determined for various probe frequencies ν and delay times t_D . The isotropic absorption signal, defined as $\Delta OD(\nu, t_D)_{\text{iso}} = -[\log(T/T_0)_\parallel + 2\log(T/T_0)_\perp]/3$ is determined and plotted as $\Delta OD \times 10^3$ in Figs 2–4, that is, independently of a possibly induced optical anisotropy.

Preparation of the investigated ice crystals. The samples are produced by slowly cooling a mixture of HDO (where D is deuterated H) in D₂O or in H₂O (15 M) between two CaF₂ windows with a suitable spacer (2.5 or $1.0 \mu\text{m}$) in a cryostat to 180 K and adjusting the desired temperature later on. The isotopic mixtures are prepared from appropriate amounts of tri-distilled H₂O and D₂O ($>99.9 \text{ at.}\%$ D) without further purification. The melting points are measured to be $274.8 \pm 0.3 \text{ K}$ and $273.4 \pm 0.3 \text{ K}$ for HDO:D₂O and HDO:H₂O, respectively.

Isochoric temperature jump. To calibrate our ‘picosecond thermometer’ we performed careful steady-state measurements of the (isobaric) temperature change of the absorption spectrum using Fourier transform infrared (FTIR) spectroscopy and also our subpicosecond infrared spectrometer (with the pump pulse blocked). In addition we took into account that the ultrafast temperature rise of the sample occurs at constant volume leading to a simultaneous pressure increase, because thermal volume expansion is relatively slow on a nanosecond timescale only. The simultaneous pressure change in the crystal during a fast temperature rise at constant volume is incorporated by using the reported isothermal pressure shift of the vibrational band and thermodynamic parameters of ice. The isobaric temperature coefficients of the OH-band position of HDO:D₂O are measured to be $(\Delta\nu_{\text{OH}}/\Delta T)_p = 0.28 \pm 0.01 \text{ cm}^{-1} \text{ K}^{-1}$ at 200 K and $0.41 \pm 0.01 \text{ cm}^{-1} \text{ K}^{-1}$ at 270 K. The required thermodynamic parameters are the coefficient of thermal volume expansion, 155 MK^{-1} , and the isothermal compressibility of 0.12 GPa^{-1} (refs 27 and 28), from which the isochoric pressure increase of 1.3 MPa K^{-1} is derived. Using the isothermal pressure downshift^{29,30} of $-78 \pm 7 \text{ cm}^{-1} \text{ GPa}^{-1}$ for the OH band of ice I_h (ref. 29) we estimate the isochoric shifts to be $(\Delta\nu_{\text{OH}}/\Delta T)_V = 0.18 \pm 0.02 \text{ cm}^{-1} \text{ K}^{-1}$ at 200 K and $0.31 \pm 0.02 \text{ cm}^{-1} \text{ K}^{-1}$ at 270 K, respectively. The OD-stretching mode is expected to display a corresponding shift approximately a factor of $\sqrt{2}$ smaller. The computed results for the thermal differential spectra for isochoric temperature jumps of 15 K (green line), 20 K (black) and 25 K (cyan) are shown in Fig. 2b. From comparison with additional curves we deduce an accuracy of our picosecond thermometer of $\pm 2 \text{ K}$ in this sample.

We have verified experimentally that the temperature dependence of the broader OH-stretching band of H₂O can be applied in an analogous manner for a picosecond thermometer. The measuring accuracy in the HDO:H₂O (15 M) ice sample is somewhat reduced to $\pm 5 \text{ K}$.

Received 11 May; accepted 3 November 2005.

1. Luo, S. N. *et al.* Maximum superheating and undercooling: systematics, molecular dynamics simulations, and dynamic experiments. *Phys. Rev. B* **68**, 134206 (2003).

2. Luo, S. N. & Ahrens, T. J. Shock-induced superheating and melting curves of geophysically important minerals. *Phys. Earth Planet. Inter.* **143–144**, 369–386 (2004).
3. Huang, K. C., Wang, T. & Joannopoulos, J. D. Superheating and induced melting at semiconductor interface. *Phys. Rev. Lett.* **94**, 175702 (2005).
4. Jin, Z. H., Gumbsch, P., Lu, K. & Ma, E. Melting mechanisms at the limit of superheating. *Phys. Rev. Lett.* **87**, 055703 (2001).
5. Cahn, R. W. Melting and the surface. *Nature* **323**, 668–669 (1986).
6. Mishima, O. & Stanley, H. E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1998).
7. Wernet, Ph. *et al.* The structure of the first coordination shell in liquid water. *Science* **304**, 995–999 (2004).
8. Torre, R., Bartolini, P. & Righini, R. Structural relaxation in supercooled water by time-resolved spectroscopy. *Nature* **428**, 296–298 (2004).
9. Ludwig, R. Water: from clusters to the bulk. *Angew. Chem. Int. Edn Engl.* **40**, 1808–1827 (2001).
10. Tse, J. S. *et al.* The mechanisms for pressure-induced amorphization of ice I_h. *Nature* **400**, 647–649 (1999).
11. Tulk, C. A. *et al.* Structural studies of several distinct metastable forms of amorphous ice. *Science* **297**, 1320–1323 (2002).
12. Mishima, O., Calvert, L. D. & Whalley, E. 'Melting ice' I at 77 K and 10 kbar: a new method of making amorphous solids. *Nature* **310**, 393–395 (1984).
13. Engemann, S. *et al.* Interfacial melting of ice in contact with SiO₂. *Phys. Rev. Lett.* **92**, 205701 (2004).
14. Ruan, Ch.-Y., Lobastov, V. A., Vigliotti, F., Chen, S. & Zewail, A. H. Ultrafast electron crystallography of interfacial water. *Science* **304**, 80–84 (2004).
15. Luo, S. N., Strachan, A. & Swift, D. C. Deducing solid-liquid interfacial energy from superheating or supercooling: application to H₂O at high pressures. *Model. Simul. Mater. Sci. Eng.* **13**, 321–328 (2005).
16. McBride, C., Vega, C., Sanz, E., MacDowell, L. G. & Abascal, J. L. F. The range of metastability of ice-water melting for two simple models of water. *Mol. Phys.* **103**, 1–5 (2005).
17. Laenen, R., Simeonidis, K. & Laubereau, A. Generation of sub-ps IR pulses via parametric processes and application to transient spectroscopy of molecules. *Laser Phys.* **9**, 234–238 (1999).
18. Woutersen, S., Emmerichs, U., Nienhuys, H. K. & Bakker, H. J. Anomalous temperature dependence of vibrational lifetimes in water and ice. *Phys. Rev. Lett.* **81**, 1106–1109 (1998).
19. Graener, H., Seifert, G. & Laubereau, A. New spectroscopy of water using tunable picosecond pulses in the infrared. *Phys. Rev. Lett.* **66**, 2092–2095 (1991).
20. Laenen, R., Rauscher, C. & Laubereau, A. Dynamics of local substructures in water observed by ultrafast infrared hole burning. *Phys. Rev. Lett.* **80**, 2622–2625 (1998).
21. Woutersen, S., Emmerichs, U. & Bakker, H. J. Femtosecond Mid-IR pump-probe spectroscopy of liquid water: evidence for a two-component structure. *Science* **278**, 658–660 (1997).
22. Fecht, H. J. Defect-induced melting and solid-state amorphization. *Nature* **356**, 133–135 (1992).
23. Górecki, T. Comments on vacancies and melting. *Scr. Metall.* **11**, 1051–1053 (1977).
24. Grishina, N. & Buch, V. Structure and dynamics of orientational defects on ice I. *J. Chem. Phys.* **120**, 5217–5225 (2004).
25. Donadio, D., Raiteri, P. & Parrinello, M. Topological defects and bulk melting of hexagonal ice. *J. Phys. Chem. B* **109**, 5421–5424 (2005).
26. Laenen, R., Simeonidis, K. & Laubereau, A. Subpicosecond spectroscopy of liquid water in the infrared: effect of deuteration on the structural and vibrational dynamics. *J. Phys. Chem. B* **106**, 408–417 (2002).
27. LaPlaca, S. & Post, B. Thermal expansion of ice. *Acta Crystallogr.* **13**, 503–505 (1960).
28. Grow, A. J. & Williamson, T. C. Linear compressibility of ice. *J. Geophys. Res.* **77**, 6348–6352 (1972).
29. Minceva-Sukarova, B., Sherman, W. F. & Wilkinson, G. R. The Raman spectra of ice (I_h, II, III, V, VI and IX) as functions of pressure and temperature. *J. Phys. C* **17**, 5833–5850 (1984).
30. Sivakumar, T. C., Chew, H. A. M. & Johari, G. P. Effect of pressure on the Raman spectrum of ice. *Nature* **275**, 524–525 (1978).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.I. (higlev@ph.tum.de).

Methane emissions from terrestrial plants under aerobic conditions

Frank Keppler¹, John T. G. Hamilton², Marc Braß^{1,3} & Thomas Röckmann^{1,3}

Methane is an important greenhouse gas and its atmospheric concentration has almost tripled since pre-industrial times^{1,2}. It plays a central role in atmospheric oxidation chemistry and affects stratospheric ozone and water vapour levels. Most of the methane from natural sources in Earth's atmosphere is thought to originate from biological processes in anoxic environments². Here we demonstrate using stable carbon isotopes that methane is readily formed *in situ* in terrestrial plants under oxic conditions by a hitherto unrecognized process. Significant methane emissions from both intact plants and detached leaves were observed during incubation experiments in the laboratory and in the field. If our measurements are typical for short-lived biomass and scaled on a global basis, we estimate a methane source strength of 62–236 Tg yr⁻¹ for living plants and 1–7 Tg yr⁻¹ for plant litter (1 Tg = 10¹² g). We suggest that this newly identified source may have important implications for the global methane budget and may call for a reconsideration of the role of natural methane sources in past climate change.

Methane (CH₄) is the most abundant organic trace gas in the atmosphere (mixing ratio ~1.8 p.p.m.) and is important to both tropospheric and stratospheric chemistry. Therefore, the atmospheric CH₄ budget has been intensively studied over the past two decades using flux measurements on sources³, global observation networks⁴ and global atmospheric models^{5,6}. In addition, stable carbon isotope ratios (¹³C/¹²C) have been applied to investigate sources and sinks of atmospheric CH₄ (refs 7, 8). Although uncertainties in the estimates of individual source strengths are large (50–100 Tg), it is generally thought that all major sources, including wetlands, animals, rice cultivation, biomass burning and fossil fuel production, have been identified and sum up to a global source strength of ~600 Tg yr⁻¹ (refs 1, 2). However, significantly elevated CH₄ mixing ratios were recently observed in tropical regions above evergreen forests⁹ indicating an additional tropical source of 30–40 Tg over the time period of the investigation (August–November), which could not be explained within the currently accepted global budget of CH₄.

Following our observations of non-enzymic production of methyl halides from senescent plants and leaf litter^{10,11}, we investigated the possibility of methane formation by plant material. A large set of laboratory experiments using freshly collected and dried plant material—including tree and grass leaves from C₃ and C₄ plant categories—were conducted, in which CH₄ release rates and stable carbon isotope composition (^δ¹³C values) of emissions were measured under controlled conditions (see Methods). Whereas CH₄ emissions were difficult to quantify for samples incubated in ambient air owing to the high atmospheric background levels of CH₄, production was clearly evident when samples were incubated in CH₄-free air. Emission rates typically ranged from 0.2 to 3 ng per g

(dry weight) h⁻¹ at 30 °C (see Supplementary Table S1). Release of CH₄ was very temperature sensitive—concentrations approximately doubled with every 10 °C increase over the range 30–70 °C (Fig. 1a), suggesting a non-enzymic rather than an enzyme-mediated process. ^δ¹³C of the emitted CH₄ ranged from –51.8‰ to –68.4‰ (mean = –58.2‰, *n* = 61) and –46.9‰ to –53.1‰ (mean = –49.5‰, *n* = 13) for C₃ and C₄ plants, respectively. The mean value determined for C₃ plant emissions is comparable with the average ^δ¹³C value for CH₄ emitted from wetlands and rice paddies (approximately –60‰; ref. 8) and thus would be generally regarded as an indication for biological production by anaerobic bacteria. Even though this possibility was remote since most of our experiments were performed under aerobic conditions, we measured CH₄ production by leaf tissue sterilized with γ-radiation (Fig. 1b, c). No significant difference, either in emission rates or ^δ¹³C values of the emissions, was noted between sterilized and non-sterilized samples, thus further excluding microbial activity as the CH₄ source and clearly indicating the existence of a hitherto unknown pathway for CH₄ production in leaf tissue.

Having established CH₄ production by detached leaf tissue, we investigated the possibility of CH₄ formation by intact plants, using incubation chambers in the laboratory and in the field (see Methods). CH₄ formation was observed for all plant species investigated, with release rates ranging from 12 to 370 ng per g (dry weight) h⁻¹, thus one to two orders of magnitude higher than the emissions from detached leaf material. Furthermore, emission rates were found to increase dramatically, by a factor of 3–5 (up to 870 ng per g (dry weight) h⁻¹), when chambers were exposed to natural sunlight, an effect also observed with detached leaf tissue (see Supplementary Information). As can be seen from Fig. 2, CH₄ concentrations increased continuously when plants were incubated in chambers at ambient temperatures. We conducted most laboratory chamber experiments in a CH₄-free air atmosphere where, in addition to concentration measurements, reliable ^δ¹³C measurements were also recorded when CH₄ concentrations in the chamber were above 40 p.p.b. (Fig. 2b–d). ^δ¹³C values were in the range of –48‰ to –59.5‰ (mean = –52‰, *n* = 29) and –45 to –47‰ (mean = –46.5‰, *n* = 11) for C₃ and C₄ plants, respectively. Our experiments were performed in a well circulated atmosphere containing ~20% oxygen, so it was unlikely that the observed CH₄ production could have been mediated by anaerobic acetate fermentation or CO₂ reduction, since obligate anaerobes metabolize only under anoxic conditions at redox levels *E*_h < –200 mV.

Nevertheless, we conducted a series of experiments that have enabled us to unequivocally demonstrate *in situ* formation of CH₄ in plants. First, we could not detect any differences in either the emission rates or the ^δ¹³C values for CH₄ produced by plants of the same species that were grown hydroponically or on soil. Second, since

¹Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, 69117 Heidelberg, Germany. ²Department of Agriculture and Rural Development for Northern Ireland, Agriculture, Food and Environmental Science Division, Newforge Lane, Belfast BT9 5PX, UK. ³Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Princetonplein 5, 3508 TA, Utrecht, The Netherlands.

the methyl group of acetate is considered to be the principal substrate from which CH_4 is readily formed by microbial activity in anaerobic soils, $2\text{-}^{13}\text{C}$ labelled acetate was added to the soil of some plants and the CH_4 released monitored for ^{13}C enrichment. Using this approach, any contribution to CH_4 production via the acetate fermentation pathway would be clearly identifiable by a massive increase in its ^{13}C content. No significant incorporation of ^{13}C was observed, eliminating microbial formation via the anoxic pathway. Finally, the difference between the $\delta^{13}\text{C}$ values for CH_4 emissions from C_3 and C_4 plants is similar to the difference between the $\delta^{13}\text{C}$ values for the biomass of the two plant categories (Supplementary Fig. S1). All these findings lead us to conclude that the observed CH_4 emissions during these studies must have originated from a thus far unknown process in the plant itself, and that this process is clearly distinct from the widely accepted process requiring anoxic conditions.

The observed ^{13}C depletion of the CH_4 emitted relative to bulk biomass indicates that the process must either have an associated large kinetic isotope effect or the source substrate must be isotopically depleted. It was recently demonstrated that the major plant C_1 (one-carbon unit) pool, which includes methoxyl groups from pectin and lignin, has a unique carbon isotope signature exceptionally depleted in ^{13}C (ref. 10). Since this pool has been shown to be responsible for the emissions of C_1 compounds such as CH_3Cl and CH_3OH from senescent leaves and leaf litter^{11,12}, its involvement in the *in situ* formation of CH_4 by plants can also be envisaged. Indeed, in experiments with purified apple pectin we also found formation of CH_4 ; emission rates, $\delta^{13}\text{C}$ values and response to temperature and sunlight were similar to those observed with detached leaves (see Supplementary Table S1 and Supplementary Fig. S2). These results

indicate that the structural plant component pectin plays a prominent role in the *in situ* formation of CH_4 in plants. However, since no chemical mechanism for CH_4 production in plants is known, an explanation for its unprecedented release must await much more detailed investigations.

To assess the global environmental impact of this newly established CH_4 source, we assume that the range of measured emission rates are globally representative for short-lived biomass. We also make the assumption that these emissions can be scaled relative to the annual net primary production (NPP), where we distinguish between the various types of biome¹³, differences in the length of vegetation period and average daily sunshine hours (Table 1). We are aware that this simple approach neglects the complexity of terrestrial ecosystems, in particular that the influence of solar radiation will be more variable than that indicated. Hence, our calculations should be considered as a first estimate. CH_4 released by living vegetation is calculated to be in the range $62\text{--}236\text{ Tg yr}^{-1}$ (average 149 Tg yr^{-1}) with the main contribution, $46\text{--}169\text{ Tg yr}^{-1}$ (average 107 Tg yr^{-1}), assigned to tropical forests and grasslands (Table 1). Between one and two orders of magnitudes lower, in the range $0.5\text{--}6.6\text{ Tg yr}^{-1}$, is our estimate for annual global production of CH_4 by plant litter.

The detection of an additional source of this magnitude, some 10–30% of the present annual source strength, would necessitate reconsideration of the global CH_4 budget. Even though in some budget estimates it is possible to accommodate an additional source of some 50 Tg in order to match the well-established sink strength², it appears that a reduction in other source terms would be necessary to keep the budget balanced. We suggest that production due to the source we have identified here may overlap with production assigned

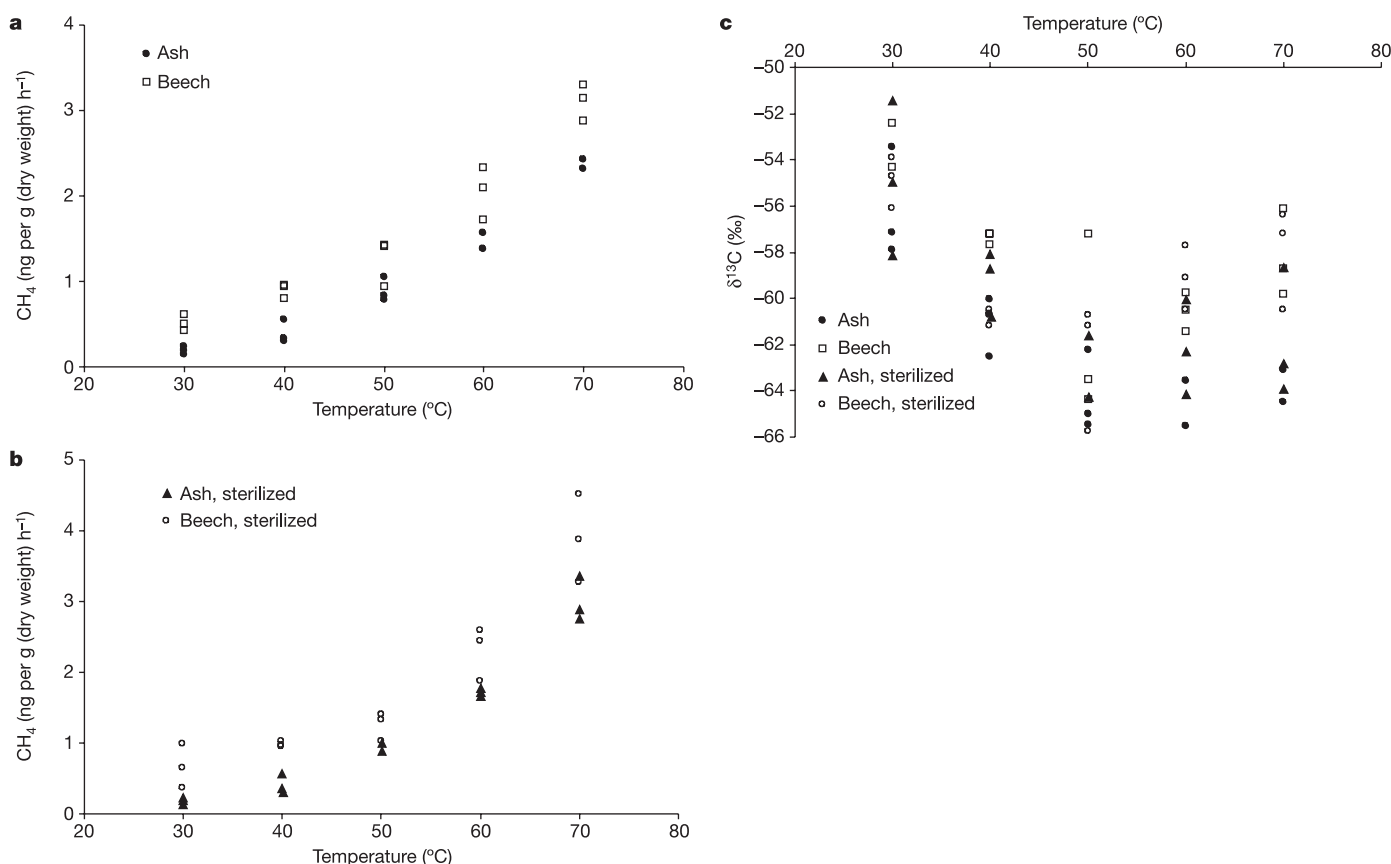


Figure 1 | Release rates and isotopic signatures of CH_4 formed by leaf tissue incubated in the dark. **a**, CH_4 release rates of air dried ash (*Fraxinus excelsior*) and beech leaves (*Fagus sylvatica*) in the temperature range

30–70 °C. **b**, CH_4 release rates (**b**) and $\delta^{13}\text{CH}_4$ values (**c**) for ash and beech leaves with and without sterilization using γ -radiation. Data from tropical plant species are shown in Supplementary Fig. S3.

to other sources such as wetlands or rice cultivation. For example, recent work with different rice cultivars showed a strong positive correlation of the crop growth parameters leaf number and leaf area index with total CH_4 flux¹⁴. In other studies with rice plants, it was shown that a 1% reduction in the solar radiation resulted in a ~2% reduction in CH_4 emissions¹⁵. A straightforward explanation for part of these emissions would be *in situ* formation in the plants themselves and, as we would anticipate, a concomitant enhancement of emissions with increased leaf biomass and solar radiation.

Direct CH_4 emissions from plants may also provide simple explanations for several findings reported in the literature that up to now were not understood. For example, the source is of the correct magnitude and also at the correct location to explain the observed elevated CH_4 levels over tropical evergreen forests⁹, a finding that could not be accounted for within the previously accepted global budget. Furthermore, since severe anthropogenic deforestation has considerably reduced tropical biomass over the past decades (a -12.3% net change in tropical forest area between 1990 and 2000; ref. 16), a corresponding reduction in tropical plant CH_4 emissions would have resulted in a decrease of 6–20 Tg yr^{-1} over this time period. This is similar in magnitude to the present global source-sink imbalance², and thus reduced biomass has probably

contributed to the recent decrease in the atmospheric growth rate of CH_4 concentration^{4,17}.

We also suggest that in pre-industrial times, that is, without anthropogenic emissions, the relative contribution of CH_4 to the atmosphere by direct plant emissions may have been even larger than today. This could have far reaching implications for the interpretation of atmospheric CH_4 levels and climate signals in the past. For example, variations in total global biomass over glacial cycles¹⁸ must contribute to reported differences in atmospheric CH_4 mixing ratios between glacial and interglacial periods¹⁹. Recent measurements of $\delta^{13}\text{C}$ values of atmospheric CH_4 from ice cores covering the past 2,000 years (ref. 20) provide additional observational support for a prominent role of a plant source in the pre-industrial atmosphere. Unexpectedly enriched $\delta^{13}\text{C}$ values of around -47‰ were shown to persist over the time period 0 to 1200 AD. This cannot be reconciled with a pre-industrial methane budget dominated by isotopically depleted wetland emissions ($\delta^{13}\text{C} \approx -60$ ‰; ref. 8), as this would lead to atmospheric $\delta^{13}\text{C}$ values in the region of -54 to -49‰ (refs 21, 22). Since direct plant emissions are enriched in ^{13}C compared to wetland emissions (from our measurements we derive a $\delta^{13}\text{C}$ value of about -50‰ based on a 60:40 ratio of C_3 and C_4 plants), the isotope mass balance for the pre-industrial atmosphere

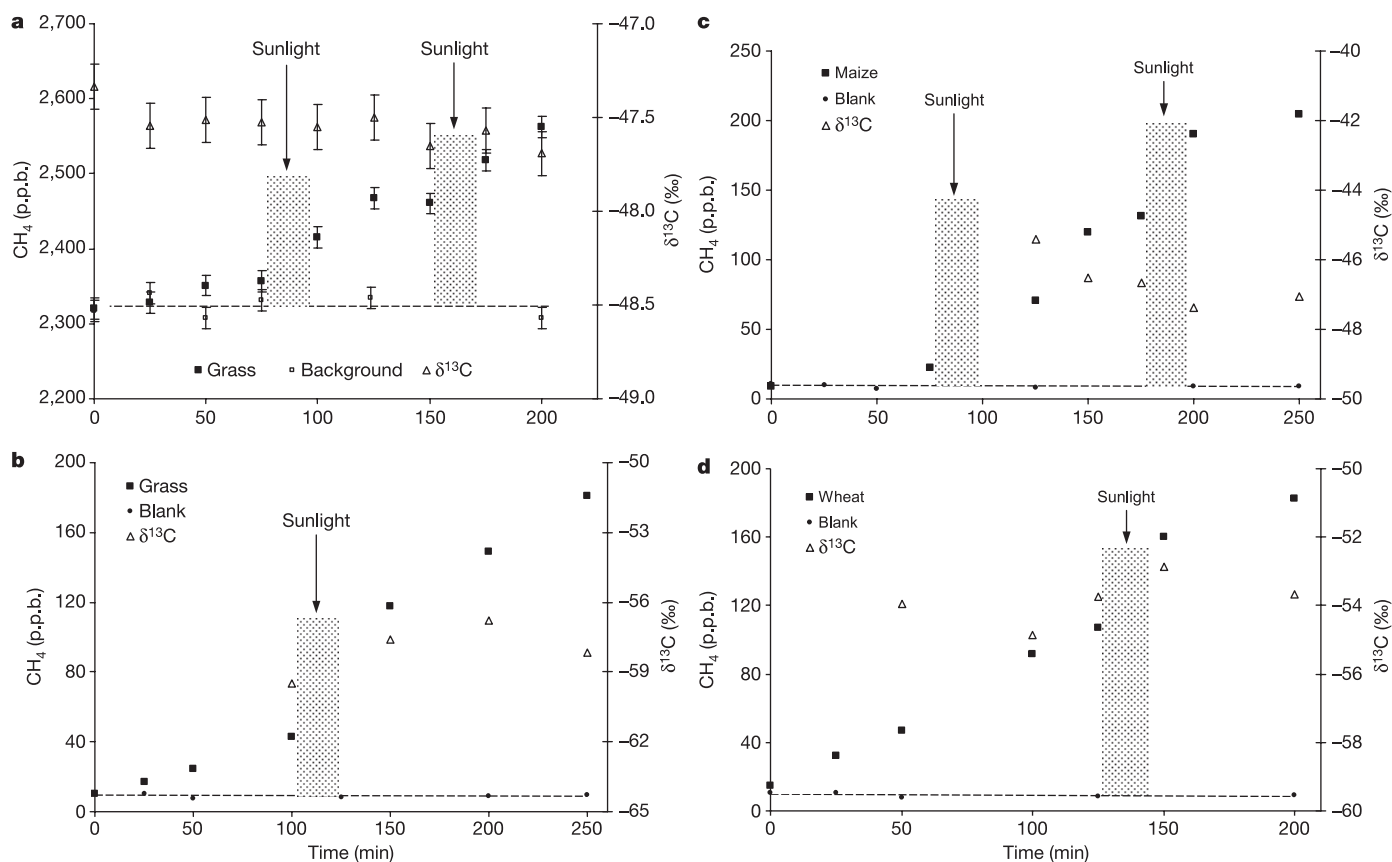


Figure 2 | Mixing ratios and $\delta^{13}\text{C}$ values of CH_4 formed by intact plants.

a, Increase of CH_4 mixing ratio during incubation of sweet vernal grass (*Anthoxanthum odoratum* L.) in ambient air. CH_4 mixing ratios in the chamber increased from 2,320 to 2,357 p.p.b. within 75 min, yielding emission rates in the range 36–126 ng per g (dry weight) h^{-1} for the individual time steps. Exposure to sunlight (indicated by dotted areas) significantly increased CH_4 release to ~320 ng per g (dry weight) h^{-1} . After sunlight exposure, release rates decreased again. A comparison between plants with and without exposure to sunlight is shown in Supplementary Fig. S4. Dashed line shows the mean CH_4 concentration without plants. $\delta^{13}\text{C}$ values of CH_4 from incubation experiments ranged from -47.3‰ to -47.7‰. Error bars shown reflect the uncertainty ($\pm 1\sigma$) of the measurements of the analytical system. **b**, Profile of CH_4 mixing ratio and

associated $\delta^{13}\text{C}$ values during incubation of sweet vernal grass (*Anthoxanthum odoratum* L., C_3 plant) in a chamber purged with CH_4 -free air before experiment. CH_4 increased from 10 to 43 p.p.b. within the first 100 min, corresponding to emission rates in the range of 48 to 63 ng per g (dry weight) h^{-1} . Sunlight exposure increased CH_4 release to ~256 ng per g (dry weight) h^{-1} . After sunlight exposure, release rates decreased to ~110 ng per g (dry weight) h^{-1} . $\delta^{13}\text{C}$ values of emitted CH_4 ranged from -56.8‰ to -59.5‰. **c**, **d**, Profiles of CH_4 concentrations and associated $\delta^{13}\text{C}$ values during incubation of maize (*Zea mays*; C_4 plant) (**c**) and wheat (*Triticum aestivum* L.; C_3 plant) (**d**) in chambers purged with CH_4 -free air before experiment. The $\delta^{13}\text{C}$ of CH_4 from maize was in the range -45.5‰ to -47.4‰, while that from wheat was more depleted in ^{13}C , from -52.7‰ to -54.9‰.

Table 1 | Estimated annual global emissions of CH₄ by living plants and leaf litter

Vegetation type/biome	Season length* (d)	NPP† (Pg C yr ⁻¹)	Sunshine hours‡ (h d ⁻¹)	Annual CH ₄ production§ low/mean/high (Tgyr ⁻¹)
Living biomass				
Tropical forests	365	21.9	8	33.2/78.2/123
Temperate forests	250	8.1	6	7.1/17.7/ 28.4
Boreal forests	150	2.6	4	1.1/3/4.1
Mediterranean shrublands	200	1.4	8	1.2/2.7/4.3
Tropical savannas and grasslands	200	14.9	8	12.4/29.2/45.9
Temperate grasslands	150	5.6	6	2.9/7.4/11.8
Deserts	100	3.5	10	1.7/3.8/5.9
Crops	200	4.1	8	2.9/7.2/11.5
Total		62.1		62.3/149/236
	Period# (d)	NPP† (Pg C yr ⁻¹)	Sunshine hours (h d ⁻¹)	
Leaf litter¶				
Tropical forests	365	21.9	8	0.23/1.53/ 3.2
Temperate forests	90	8.1	6	0.02/0.12/0.25
Boreal forests	60	2.6	4	0.01/0.02/0.05
Mediterranean shrublands	180	1.4	8	0.01/0.05/0.1
Tropical savannas and grasslands	365	14.9	8	0.16/ 1/2.1
Temperate grasslands	90	5.6	6	0.01/0.08/0.18
Deserts	365	3.5	10	0.04/0.28/0.56
Crops	90	4.1	8	0.01/0.07/0.15
Total		62.1		0.49/3.2/6.6

* Estimated.

† Data from ref. 13.

‡ Estimated hours of sunshine per day during vegetation period.

§ For calculation, see Methods.

|| Low (high) estimates are derived as mean of the CH₄ emission rates – (+) 1σ, value for the mean is similar to the median.

¶ Emissions measured by detached leaf tissue (fresh and dried) are considered to reflect those from leaf litter.

Estimated period of plant decay with similar ambient temperatures as in our experiments.

can be closed if plant emissions are included as an important natural CH₄ source. Consequently, the role of natural CH₄ sources in past climate change, particularly when biospheric productivity changed dramatically, may have to be reconsidered.

Finally, it has been suggested that the so-called 'CO₂ fertilization effect' could lead to a substantial increase in NPP over the next 100 years (ref. 23), which should have an effect on CH₄ emissions from plants. Thus for the future it is essential that we fully understand the relationship between climate change and plant CH₄ emissions.

METHODS

Incubation experiments with detached leaves. Fresh leaves (1–6 g; detached from the intact plants) and dried leaves (1–5 g, air dried at 25 °C for 48 h) were placed in glass vials (44 ml) and sealed with caps containing PTFE-lined silicon septa. Vials were purged with CH₄-free air for one hour before the start of the experiment, and controls (blanks) were measured after purging. After incubation in the dark for 16 h at 30 °C and 40 °C, CH₄ formed in the vial was analysed with continuous-flow isotope ratio mass spectrometry (CF-IRMS). Leaf dry matter was determined by drying at 105 °C for 24 h. Experiments with solar radiation were performed by placing glass vials in direct sunlight for a 1 h period between 10:00 h and 15:00 h in Heidelberg, Germany.

Chamber experiments with intact plants. Plexiglas chambers (volume 18 l, diameter $d = 29$ cm) were used for static incubation experiments. Plants were placed in sealed chambers and purged with CH₄-free air until CH₄ background levels were below 10 p.p.b. Chamber temperature, pressure, CO₂ concentration and humidity and external atmospheric pressure were all monitored online during the experimental period. In all experiments, CO₂ concentration never dropped below 300–250 p.p.m. Concentrations and $\delta^{13}\text{C}$ values of CH₄ were measured every 25 min by transferring 40 ml of chamber headspace gas into the analytical system. Small electric fans were used to circulate the chamber air. Dry matter content was determined at the end of the experiment. Experiments with solar radiation were performed by placing chambers in direct sunlight for a 20 min period between 10:00 h and 15:00 h in Heidelberg, Germany.

Methane measurements. Headspace gas samples were transferred from the sample vial or the incubation chamber to an evacuated 40 cm³ sample loop. CH₄ was trapped on Haysep D, separated by gas chromatography from interfering compounds and transferred via an open split to the isotope ratio mass spectrometer (ThermoFinnigan Delta^{plus} XL). Concentration and $\delta^{13}\text{C}$ values were determined using an air standard with known concentration and isotopic

composition as internal reference, and a measurement of the inlet pressure in the sample loop. The reference gas was also used to establish a robust linearity correction, since sample CH₄ concentrations were very variable. $\delta^{13}\text{C}$ values are reported relative to Vienna-PDB and defined by the equation $\delta^{13}\text{C} = (R_{\text{sample}}/R_{\text{V-PDB}} - 1) \times 1,000\text{‰}$, with $R = {}^{13}\text{C}/{}^{12}\text{C}$.

Calculation of annual CH₄ production. As a first estimate, average daily CH₄ emission rates are calculated as $\text{ER}_{\text{day}} = (\text{ER}_{\text{sun}} \times h_{\text{sun}}) + (\text{ER}_{\text{nosun}} \times h_{\text{nosun}})$, where ER_{sun} and ER_{nosun} are the measured emission rates of CH₄ with/without direct sunshine (10⁻⁹ per g (dry weight) biomass h⁻¹), and h_{sun} and h_{nosun} are the estimated daily hours with/without sunshine (h d⁻¹), respectively.

Mean (low/high) values of ER_{sun} for intact plants and detached leaves were 374 (198/598) and 8.7 (1.6/15.8) ng per g (dry weight) h⁻¹, respectively. Mean (low/high) values of ER_{nosun} for living plants and detached leaves were 119 (30.7/207) and 1.6 (0.1/4.4) ng per g (dry weight) h⁻¹, respectively. Low (high) estimates are derived as mean of the CH₄ emission rates – (+) 1σ. Solar radiation experiments were carried out in direct sunshine in springtime (March–May 2005) in Heidelberg, Germany (~49.4°N); living plants: 8 different plant species, $n = 33$; detached leaves: 14 different plant species, $n = 23$ (see Supplementary Information). 'No sun' experiments were carried out in the laboratory without direct solar radiation; living plants: 9 different plant species, $n = 46$; detached leaves: 19 different plant species, $n = 61$ (see Supplementary Information).

Annual production of CH₄ per biome type is calculated as $P(\text{CH}_4)_{\text{annual}} = \text{NPP} \times 2 \times \text{SL} \times \text{ER}_{\text{day}}$, where $P(\text{CH}_4)_{\text{annual}}$ is the annual production (gyr⁻¹), NPP is the net primary production of the biome type (10¹⁵ g C yr⁻¹) and SL is season length (d); the factor 2 is needed to convert NPP, which is usually expressed as carbon equivalent, to plant biomass, assuming that plant biomass is 50% carbon.

Received 14 July; accepted 3 November 2005.

- Lelieveld, J., Crutzen, P. J. & Dentener, F. J. Changing concentration, lifetimes and climate forcing of atmospheric methane. *Tellus B* **50**, 128–150 (1998).
- Houghton, J. T. et al. (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, 2001).
- Bartlett, K. B. & Harriss, R. C. Review and assessment of methane emissions from wetlands. *Chemosphere* **26**, 261–320 (1993).
- Dlugokencky, E. J., Masarie, K. A., Lang, P. M. & Tans, P. P. Continuing decline in the growth rate of the atmospheric methane burden. *Nature* **393**, 447–450 (1998).
- Walter, B. P., Heimann, M. & Matthews, E. Modeling modern methane emissions from natural wetlands 2. Interannual variations 1982–1993. *J. Geophys. Res.* **106**, 34207–34219 (2001).

6. Houweling, S., Dentener, F., Lelieveld, J., Walter, B. & Dlugokencky, E. The modeling of tropospheric methane: How well can point measurements be reproduced by a global model? *J. Geophys. Res.* **105**, 8981–9002 (2000).
7. Lowe, D. C. *et al.* Concentration and ^{13}C records of atmospheric methane in New Zealand and Antarctica—Evidence for changes in methane sources. *J. Geophys. Res.* **99**, 16913–16925 (1994).
8. Quay, P. *et al.* The isotopic composition of atmospheric methane. *Glob. Biogeochem. Cycles* **13**, 445–461 (1999).
9. Frankenberg, C., Meirink, J. F., van Weele, M., Platt, U. & Wagner, T. Assessing methane emissions from global space-borne observations. *Science* **308**, 1010–1014 (2005).
10. Keppler, F., Kalin, R. M., Harper, D. B., McRoberts, W. C. & Hamilton, J. T. G. Dramatic ^{13}C depletion in the plant methoxyl pool and its biogeochemical implications. *Biogeosciences* **1**, 123–131 (2004).
11. Hamilton, J. T. G., McRoberts, W. C., Keppler, F., Kalin, R. M. & Harper, D. B. Chloride methylation by plant pectin: An efficient environmentally significant process. *Science* **301**, 206–209 (2003).
12. Keppler, F., Harper, D. B., Röckmann, T., Moore, R. M. & Hamilton, J. T. G. New insight into the atmospheric chloromethane budget gained using stable carbon isotope ratios. *Atmos. Chem. Phys.* **5**, 2403–2411 (2005).
13. Saugier, B., Roy, J. & Mooney, H. A. in *Global Terrestrial Productivity* (eds Roy, J., Saugier, B. & Mooney, H. A.) 541–555 (Academic, San Diego, 2001).
14. Gogoi, N., Baruah, K. K., Gogoi, B. & Gupta, P. K. Methane emission characteristics and its relations with plant and soil parameters under irrigated rice ecosystem of northeast India. *Chemosphere* **59**, 1677–1684 (2005).
15. Sass, R. L. & Cicerone, R. J. Photosynthate allocations in rice plants: Food production or atmospheric methane? *Proc. Natl Acad. Sci. USA* **99**, 11993–11995 (2002).
16. FAO, *Global Forest Resources Assessment 2000—Main Report* (FAO Forestry Paper No. 140, Food and Agriculture Organization of the United Nations, Rome, 2001).
17. Dlugokencky, E. J. *et al.* Atmospheric methane levels off: Temporary pause or a new steady-state? *Geophys. Res. Lett.* **30** doi:10.1029/2003GL018126 (2003).
18. Mayle, F. E., Beerling, D. J., Gosling, W. D. & Bush, M. B. Responses of Amazonian ecosystems to climatic and atmospheric carbon dioxide changes since the last glacial maximum. *Phil. Trans. R. Soc. Biol. Sci.* **359**, 499–514 (2004).
19. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
20. Ferretti, D. *et al.* Unexpected changes to the global methane budget over the past 2000 years. *Science* **309**, 1714–1717 (2005).
21. Lassey, K. R., Lowe, D. C. & Manning, M. R. The trend in atmospheric methane $\delta^{13}\text{C}$ implications for isotopic constraints on the global methane budget. *Glob. Biogeochem. Cycles* **14**, 41–49 (2000).
22. Houweling, S., Dentener, F. & Lelieveld, J. Simulation of preindustrial atmospheric methane to constrain the global source strength of natural wetlands. *J. Geophys. Res.* **105**, 17243–17255 (2000).
23. Cramer, W. *et al.* Global response of terrestrial ecosystem structure and function to CO_2 and climate change: results from six dynamic global vegetation models. *Glob. Change Biol.* **7**, 357–373 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Conrad, J. Kesselmeier and D. Harper for comments on the manuscript; B. Knape, P. Franz, R. Shaheen, F. Kleinbongardt, V. Mallinger, R. Runck and C. McRoberts for technical assistance; the Botanical Garden of the University of Heidelberg for providing plant species from tropical regions; and the European Commission for a Marie Curie-Research Training Grant (F.K.). The ISOSTRAT project in Heidelberg was funded by the BMBF within the AFO2000 project.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.K. (frank.keppler@mpi-hd.mpg.de).

LETTERS

Stability of hydrous melt at the base of the Earth's upper mantle

Tatsuya Sakamaki¹, Akio Suzuki¹ & Eiji Ohtani¹

Seismological observations have revealed the existence of low-velocity and high-attenuation zones above the discontinuity at 410 km depth, at the base of the Earth's upper mantle^{1,2}. It has been suggested that a small amount of melt could be responsible for such anomalies^{3–5}. The density of silicate melt under dry conditions has been measured at high pressure^{4,6–13} and found to be denser than the surrounding solid, thereby allowing the melt to remain at depth. But no experimental investigation of the density of hydrous melt has yet been carried out. Here we present data constraining the density of hydrous basaltic melt under pressure to examine the stability of melt above the 410-km discontinuity. We infer that hydrous magma formed by partial melting above the 410-km discontinuity may indeed be gravitationally stable, thereby supporting the idea that low-velocity or high-attenuation regions just above the mantle transition zone may result from the presence of melt.

The density of silicate melt was determined using the sink/float test, which has been widely used for dry melt^{4,7–13}. The crystalline density marker moves upward or downward depending on the density difference between the marker and the surrounding melt; thus, the density of melt is bounded by the density of the marker. Here we employed diamond as a density marker, because it is not reactive with silicate melt. The composition of hydrous MORB (mid-ocean ridge basalt), which was synthesized at 1 atm, is given in Table 1. The density of dry MORB melt has been previously reported⁸. The zero-pressure densities of the melt are calculated using the partial molar volumes of the oxide components in silicate liquid¹⁴ and the partial molar volume of H₂O (ref. 15).

Our density measurements were carried out in the pressure range from 14.0 to 20.0 GPa at 2,200 to 2,300 °C. The diamond marker was neutrally buoyant at 16.8 GPa and 2,300 °C in the hydrous MORB melt containing 2.0 wt% H₂O, and at 20.0 GPa and 2,200 °C in the hydrous MORB melt containing 8.0 wt% H₂O. The results of the experiments are summarized in Fig. 1. On the basis of our buoyancy test, the densities of the hydrous MORB melt containing 2.0 and

8.0 wt% H₂O were determined to be $3.55 \pm 0.05 \times 10^3 \text{ kg m}^{-3}$ at 16.8 GPa and 2,300 °C and $3.58 \pm 0.02 \times 10^3 \text{ kg m}^{-3}$ at 20.0 GPa and 2,200 °C, respectively, by using an equation of state of diamond^{16,17}. Assuming that the pressure derivative of the isothermal bulk modulus (K') is the same as that of the dry MORB melt, that is, $K' = 5.0 \pm 0.7$ (ref. 8), the isothermal bulk moduli (K_T) of the hydrous MORB melt containing 2.0 and 8.0 wt% H₂O were calculated (using the Birch–Murnaghan equation of state) to be $K_T = 13.8 \pm 2.2 \text{ GPa}$ and $6.5 \pm 1.6 \text{ GPa}$, respectively.

Previous melting experiments on water-saturated mantle peridotite have shown that the wet solidus is approximately 1,000 °C in the transition zone¹⁸. Under the conditions of undersaturation of water, the solidus temperature increases depending on the water content. The maximum water solubilities in olivine and wadsleyite in ascending plumes are 100–200 p.p.m. and 0.5 wt%, respectively^{5,19}. It has been suggested that the ascending plume is hydrated in the hydrous transition zone and then releases water by the wadsleyite–olivine phase transition at the base of the upper mantle⁵. Because water decreases the melting temperature, the ascending plume is partially

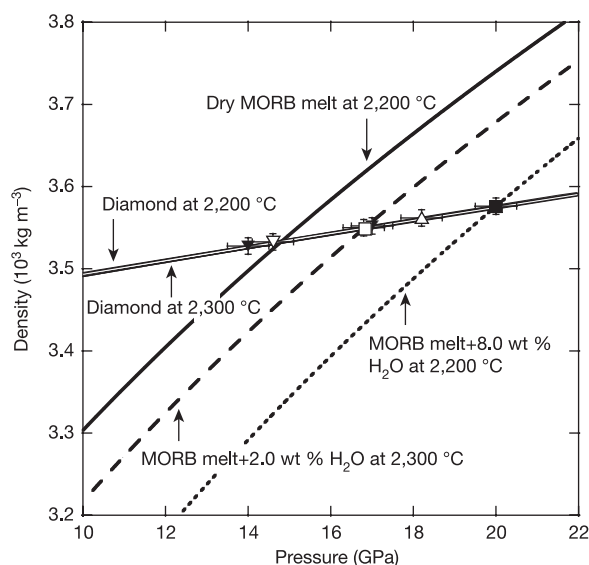


Figure 1 | Pressure versus density diagram summarizing the density measurement of hydrous MORB. Open triangle, flotation of diamond in MORB melt + 2.0 wt% H₂O. Open square, neutral buoyancy of diamond in MORB melt + 2.0 wt% H₂O. Open downtriangle, sinking of diamond in MORB melt + 2.0 wt% H₂O. Solid square, neutral buoyancy of diamond in MORB melt + 8.0 wt% H₂O. Solid downtriangles, sinking of diamond in MORB melt + 8.0 wt% H₂O. The compression curve of dry MORB (solid line) was determined previously⁸. The error bars correspond to the 95% confidence interval.

Table 1 | Compositions of MORB, hydrous MORB (2.0 and 8.0 wt% H₂O) and IT8720

	MORB	MORB + 2.0 wt% H ₂ O	MORB + 8.0 wt% H ₂ O	IT8720
SiO ₂	51.8	50.8	47.7	41.2
Al ₂ O ₃	16.0	15.6	14.7	3.7
FeO	10.0	9.8	9.2	15.1
MgO	7.9	7.7	7.2	33.0
CaO	11.7	11.5	10.8	7.0
Na ₂ O	2.7	2.7	2.5	–
H ₂ O	–	2.0	8.0	–
Total (wt%)	100.0	100.0	100.0	100.0
Mg#	58.4	58.4	58.4	79.6

Mg# = MgO/(MgO + FeO) × 100, molar.

¹Department of Earth and Planetary Material Sciences, Tohoku University, Sendai 980-8578, Japan.

molten above the transition zone. To evaluate the gravitational stability of hydrous melt formed by dehydration melting at the base of the upper mantle, we calculated the density of hydrous magma at 1,600 °C, which is the temperature of the ascending plume⁵.

Here, we evaluated the partial molar volume of water in hydrous magma by using the density of hydrous MORB. The density of anhydrous MORB melt⁸ was estimated with an accuracy of $\pm 0.02 \times 10^3 \text{ kg m}^{-3}$ at 16.8 GPa and $\pm 0.04 \times 10^3 \text{ kg m}^{-3}$ at 20.0 GPa. Subtracting the volumes of the dry melt from that of the hydrous melt for the unit mass, we obtained the excess volume of hydrous melt, that is, the partial volume of water in the hydrous MORB melt of the mass. Using that volume, we estimate that the partial molar volumes of water in hydrous MORB melt at 16.8 and 20.0 GPa and 2,200 °C are 8.5 ± 1.4 and $7.5 \pm 0.6 \text{ cm}^3 \text{ mol}^{-1}$, respectively. In this calculation, we assumed linear mixing of the basalt and H₂O components for calculating the partial molar volume of H₂O at high pressure and high temperature. We may need to evaluate carefully the applicability of assuming linear mixing at a pressure around 20 GPa, although it has been suggested that the linear-mixing model of H₂O is shown to be applicable for magmas at relatively low pressures of around 0.85 GPa (ref. 15).

The density of the NaAlSi₃O₈-H₂O system has been measured up to 0.85 GPa and 950 °C (ref. 20). The partial molar volume of water in silicate melts has been determined at 1 bar (refs 15, 21). By using these low-pressure data together with our present results, the equation of state of water species in silicate melts is determined to be $K_T = 1.3 \pm 0.9 \text{ GPa}$ and $K' = 3.7 \pm 0.2$, using the Birch–Murnaghan equation of state. The bulk modulus and its pressure derivative are $0.9 \pm 0.6 \text{ GPa}$ and 7.6 ± 1.0 , respectively, in the case of the Vinet equation of state²², which is superior to other types of equations of state for soft materials. Figure 2 shows the compression curves of the partial molar volume of H₂O in magma at 2,200 °C, which runs parallel to that of the molar volume of H₂O (ref. 23). This enables us to estimate the density of hydrous silicate melt in the Earth's interior.

Figure 3a shows the compression curves of the hydrous MORB

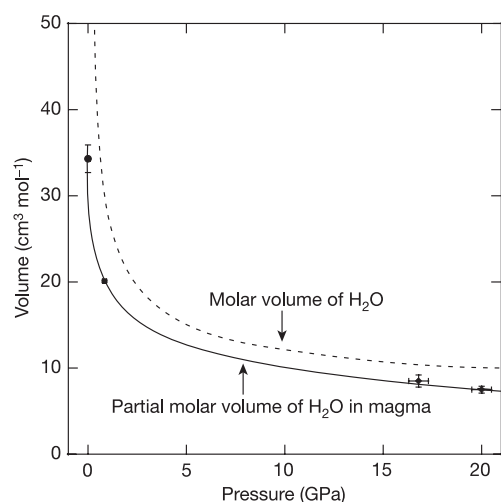


Figure 2 | Compression curves of the molar volume of H₂O (broken line) and the partial molar volume of H₂O in magma (solid line) at 2,200 °C. Partial molar volumes at 1 atm (ref. 15) and at 0.85 GPa (ref. 20) are shown as a solid circle and a solid square, respectively. Solid diamonds, data from the present study. The error bars represent the upper and lower bounds derived from the density measurement of hydrous magma by the sink/float method. The four data for the partial molar volume of H₂O are fitted by a Vinet equation of state²² with the parameters $K_T = 0.9 \pm 0.6 \text{ GPa}$ and $K' = 7.6 \pm 1.0$. Compression curve of the molar volume of H₂O (ref. 23) is shown for comparison.

melt, which was calculated from the compression curves of the partial molar volume of H₂O (Fig. 2) and dry MORB by assuming $dK_T/dT = 0$. Calculating from Fig. 3a, the density of the hydrous MORB melt with $3.0 \pm 0.3 \text{ wt\%}$ water is equal to that of PREM (Preliminary Reference Earth Model)²⁴ at 13.4 GPa and 1,600 °C. The initial melt formed by partial melting of peridotitic composition at the base of the upper mantle is ultrabasic in both dry and wet conditions^{3,25,26}. However, the chemical composition of the initial melt of water undersaturated iron-bearing peridotite has not yet been sufficiently determined. To investigate the stability of the melt, we calculated the density of the hydrous ultrabasic silicate melt. We assumed that the composition without water is similar to the partial melt formed by melting of dry peridotite, that is, IT8720 (refs 7, 27). The composition of the IT8720 melt is CaO-rich and Al₂O₃-poor

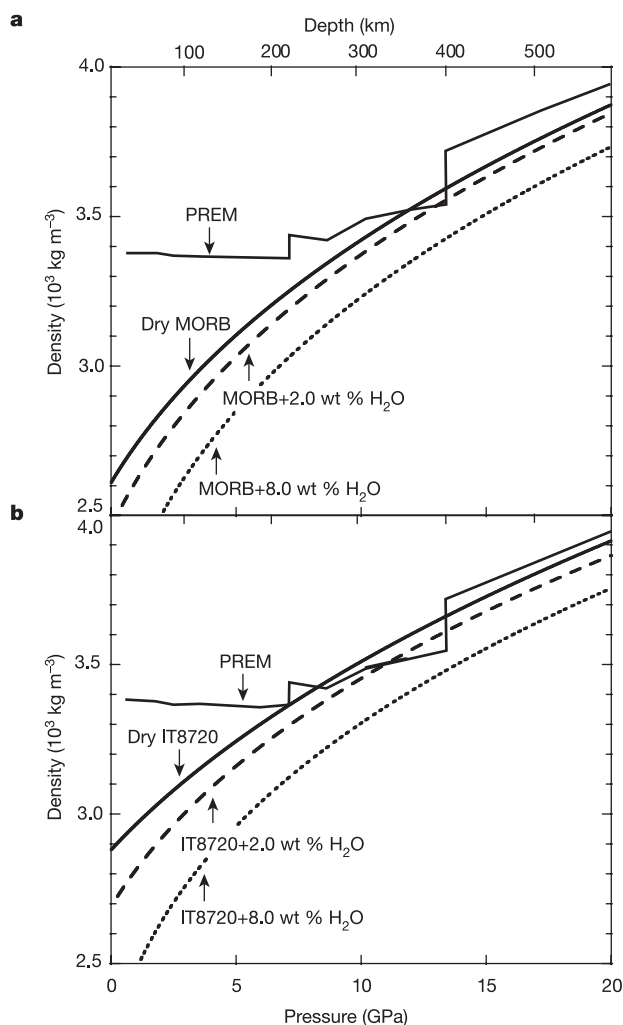


Figure 3 | Compression curves of hydrous melts at 1,600 °C and PREM (Preliminary Reference Earth Model)²⁴. **a**, The density curves of hydrous MORB melts are expressed by the Birch–Murnaghan equation of state with a zero-pressure density $\rho_0 = 2.46 \pm 0.02 \times 10^3 \text{ kg m}^{-3}$, isothermal bulk modulus $K_T = 13.9 \pm 0.1 \text{ GPa}$ and a pressure derivative $K' = 5.2 \pm 0.1$ for hydrous MORB + 2.0 wt% H₂O (broken line), and $\rho_0 = 2.09 \pm 0.02 \times 10^3 \text{ kg m}^{-3}$, isothermal bulk modulus $K_T = 6.7 \pm 0.2 \text{ GPa}$ and a pressure derivative $K' = 5.7 \pm 0.1$ for hydrous MORB + 8.0 wt% H₂O (dotted line). **b**, The density curves of hydrous peridotitic IT8720 melts are expressed by the Birch–Murnaghan equation of state with a zero-pressure density $\rho_0 = 2.69 \pm 0.01 \times 10^3 \text{ kg m}^{-3}$, bulk modulus isothermal $K_T = 19.5 \pm 0.6 \text{ GPa}$ and a pressure derivative $K' = 5.8 \pm 0.2$ for hydrous IT8720 + 2.0 wt% H₂O (broken line), and $\rho_0 = 2.24 \pm 0.01 \times 10^3 \text{ kg m}^{-3}$, bulk modulus isothermal $K_T = 6.9 \pm 0.3 \text{ GPa}$ and a pressure derivative $K' = 7.2 \pm 0.3$ for hydrous IT8720 + 8.0 wt% H₂O (dotted line).

compared to mantle peridotite and is similar to that of the partial melt of the hydrous CMAS-pyrolite at 13.5 GPa (ref. 3). Therefore, we propose that the hydrous IT8720 composition is a good candidate for the partial melt at the base of the upper mantle.

Figure 3b shows the compression curves of the dry and hydrous IT8720 melts. We have determined the K – K' relation of the dry IT8720 melt, but K' of the IT8720 melt was not determined uniquely⁷. Therefore, K' of the PHN1611 melt, which was recently determined⁴, was used to calculate K_T of the IT8720 melt, because both compositions are peridotitic. Using the bulk modulus and its pressure derivative of $K_T = 32$ GPa and $K' = 4.6$, the density of the dry IT8720 melt is $3.66 \pm 0.02 \times 10^3 \text{ kg m}^{-3}$ at 13.4 GPa and 1,600 °C. By using the partial molar volume of water determined at 16.8 and 20.0 GPa in the present study, the density of the hydrous IT8720 melt with 2.0 and 8.0 wt% H₂O are calculated to be $3.77 \pm 0.05 \times 10^3$ at 16.8 GPa and $3.80 \pm 0.04 \times 10^3 \text{ kg m}^{-3}$ at 20.0 GPa. Compression curves of the hydrous IT8720 melt in Fig. 3b shows that the melt with 6.7 ± 0.6 wt% water is estimated to be equal to that of PREM²⁴ at 13.4 GPa and 1,600 °C.

The driving force of ascending magma is the density difference between the magma and the surrounding mantle. Figure 3b shows that the density of hydrous magma is close to that of the mantle at the 410-km seismic discontinuity. The magma, which is generated in the transition zone, therefore moves upward and is decelerated at the 410-km discontinuity. The present study suggests that the hydrous magma may be gravitationally stable in the case that the water content of magma is less than 6.7 ± 0.6 wt%, and may be responsible for the seismic anomaly above the 410-km discontinuity.

METHODS

Starting materials. The 10-g mixtures of reagent-grade oxides and carbonate were ground in an agate mortar with acetone for 1 h. The mixed powder was melted in a gas-mixing furnace and the oxygen fugacity was controlled. The mixture was then quenched to make glass, and the glass was mixed with Al(OH)₃ powder in an agate mortar for 1 h. Finally, the mixed powder was heated at 130 °C for 3 h to remove the absorbed water. The water contents of the starting materials are estimated to be 2.0 ± 0.5 wt% and 8.0 ± 0.5 wt% by the measurement of the loss on ignition using 4 g of the starting material.

Experimental procedure. High-pressure experiments were carried out using a KAWAI-type multianvil apparatus²⁸ driven by a 1,000-ton uniaxial press installed at Tohoku University. The charge was pressurized to the desired pressure at room temperature, and then heated using an electric furnace. The uncertainty of the pressure is estimated to be ± 0.5 GPa. Temperatures were measured with a W3%Re–W25%Re thermocouple. No correction was made for the effect of pressure on electromotive force. The uncertainty of the temperature is estimated to be ± 25 °C. The sample was quenched by shutting off the electric power supply. Recovered samples were polished to make a thin section to examine the position of the diamond density marker. The errors on the pressure, temperature and equation of state of diamond are considered to calculate the density of melt. The water contents of the recovered samples could not be measured because they were not homogeneous glasses but consisted of dendritic quenched crystals, interstitial glasses and pores.

Received 27 April; accepted 18 October 2005.

1. Revenaugh, J. & Sipkin, S. A. Seismic evidence for silicate melt atop the 410-km mantle discontinuity. *Nature* **369**, 474–476 (1994).
2. Song, T. A., Helmberger, D. V. & Grand, S. P. Low-velocity zone atop the 410-km seismic discontinuity in the northwestern United States. *Nature* **427**, 530–533 (2004).
3. Litasov, K. & Ohtani, E. Phase relations and melt compositions in

CMAS-pyrolite–H₂O system up to 25 GPa. *Phys. Earth Planet. Inter.* **134**, 105–127 (2002).

4. Suzuki, A. & Ohtani, E. Density of peridotite melts at high pressure. *Phys. Chem. Miner.* **30**, 449–456 (2003).
5. Ohtani, E., Litasov, K., Hosoya, T., Kubo, T. & Kondo, T. Water transport into the deep mantle and formation of a hydrous transition zone. *Phys. Earth Planet. Inter.* **143–144**, 255–269 (2004).
6. Rigden, S. M., Ahrens, T. J. & Stolper, E. M. Densities of liquid silicates at high pressures. *Science* **226**, 1071–1073 (1984).
7. Suzuki, A., Ohtani, E. & Kato, T. Flotation of diamond in mantle melt at high pressure. *Science* **269**, 216–218 (1995).
8. Ohtani, E. & Maeda, M. Density of basaltic melt at high pressure and stability of the melt at the base of the lower mantle. *Earth Planet. Sci. Lett.* **193**, 69–75 (2001).
9. Agee, C. B. & Walker, D. Static compression and olivine flotation in ultrabasic silicate liquid. *J. Geophys. Res.* **93**, 3437–3449 (1988).
10. Agee, C. B. & Walker, D. Olivine flotation in mantle melt. *Earth Planet. Sci. Lett.* **114**, 315–324 (1993).
11. Agee, C. B. Crystal-liquid density inversions in terrestrial and lunar magmas. *Phys. Earth Planet. Inter.* **107**, 63–74 (1998).
12. Ohtani, E., Suzuki, A. & Kato, T. Flotation of olivine in the peridotite melt at high pressure. *Proc. Jpn Acad. Ser. B* **69**, 23–28 (1993).
13. Suzuki, A., Ohtani, E. & Kato, T. Density and thermal expansion of a peridotite melt at high pressure. *Phys. Earth Planet. Inter.* **107**, 53–61 (1998).
14. Lange, R. A. & Carmichael, I. S. E. Densities of Na₂O–K₂O–CaO–MgO–FeO–Fe₂O₃–Al₂O₃–TiO₂–SiO₂ liquids: new measurements and derived partial molar properties. *Geochim. Cosmochim. Acta* **51**, 2931–2946 (1987).
15. Ochs, F. A. III & Lange, R. A. The density of hydrous magmatic liquids. *Science* **283**, 1314–1317 (1999).
16. Zouboulis, E. S., Grimsditch, M., Randas, A. K. & Rodriguez, S. Temperature dependence of the elastic moduli of diamond: a Brillouin-scattering study. *Phys. Rev. B* **57**, 2889–2896 (1998).
17. McSkimin, H. J. & Andreatch, P. Jr Elastic moduli of diamond as a function of pressure and temperature. *J. Appl. Phys.* **43**, 2944–2948 (1972).
18. Kawamoto, T. Hydrous phase stability and partial melt chemistry in H₂O-saturated KLB-1 peridotite up to the uppermost lower mantle conditions. *Phys. Earth Planet. Inter.* **143–144**, 387–395 (2004).
19. Wallace, P. J. Water and partial melting in mantle plumes: inferences from the dissolved H₂O concentrations of Hawaiian basaltic magmas. *Geophys. Res. Lett.* **25**, 3639–3642 (1998).
20. Burnham, C. W. & Davis, N. F. The role of H₂O in silicate melts I. P–V–T relations in the system NaAlSi₃O₈–H₂O to 10 kilobars and 1000 °C. *Am. J. Sci.* **270**, 54–79 (1971).
21. Ochs, F. A. III & Lange, R. A. The partial molar volume, thermal expansivity, and compressibility of H₂O in NaAlSi₃O₈ liquid: new measurements and an internally consistent model. *Contrib. Mineral. Petrol.* **129**, 155–165 (1997).
22. Vinet, P., Rose, J. H., Ferrante, J. & Smith, J. R. Universal features of the equation of state of solids. *J. Phys. Condens. Matter* **1**, 1941–1963 (1989).
23. Belonoshko, A. & Saxena, S. K. A molecular dynamics study of the pressure-volume-temperature properties of super-critical fluids: I. H₂O. *Geochim. Cosmochim. Acta* **55**, 381–387 (1991).
24. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
25. Zhang, J. & Herzberg, C. Melting experiments on anhydrous peridotite KLB-1 from 5.0 to 22.5 GPa. *J. Geophys. Res.* **99**, 17729–17742 (1994).
26. Litasov, K. & Ohtani, E. Stability of various hydrous phases in CMAS pyrolite–H₂O system up to 25 GPa. *Phys. Chem. Miner.* **30**, 147–156 (2003).
27. Ito, E. & Takahashi, E. Melting of peridotite at uppermost lower-mantle conditions. *Nature* **328**, 514–517 (1987).
28. Kawai, N. & Endo, S. The generation of ultrahigh hydrostatic pressures by a split sphere apparatus. *Rev. Sci. Instrum.* **41**, 1178–1181 (1970).

Acknowledgements We thank T. Kondo and H. Terasaki for discussions. This work was conducted as a part of the 21st Century Center-of-Excellence Program, 'Advanced Science and Technology Center for the Dynamic Earth', at Tohoku University.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.S. (a-suzuki@mail.tains.tohoku.ac.jp).

LETTERS

A Cretaceous symmetrodont therian with some monotreme-like postcranial features

Gang Li¹ & Zhe-Xi Luo^{2,3}

A new spalacotheriid mammal preserved with a complete postcranium and a partial skull has been discovered from the Yixian Formation^{1–3} of Liaoning, China. Spalacotheroid symmetrodonts^{4–11} are relatives to modern therians (combined group of marsupials and placentals) and are characterized by many skeletal apomorphies of therians. But unlike the closely related spalacotheroids and living therians, this new mammal revealed some surprisingly convergent features to monotremes in the lumbar vertebrae, pelvis and hindlimb^{12,13}. These peculiar features may have developed as functional convergence to locomotory features of monotremes, or the presence of lumbar ribs in this newly

discovered mammal and their absence in its close relatives might be due to evolutionary developmental homoplasy. Analysis including this new taxon suggests that spalacotheroids evolved earlier in Eurasia and then dispersed to North America, in concordance with prevailing geodispersal patterns of several common mammalian groups during the Early Cretaceous period.

Class Mammalia
Clade Trechnotheria
Family Spalacotheriidae
Akidolestes cifellii gen. et sp. nov.

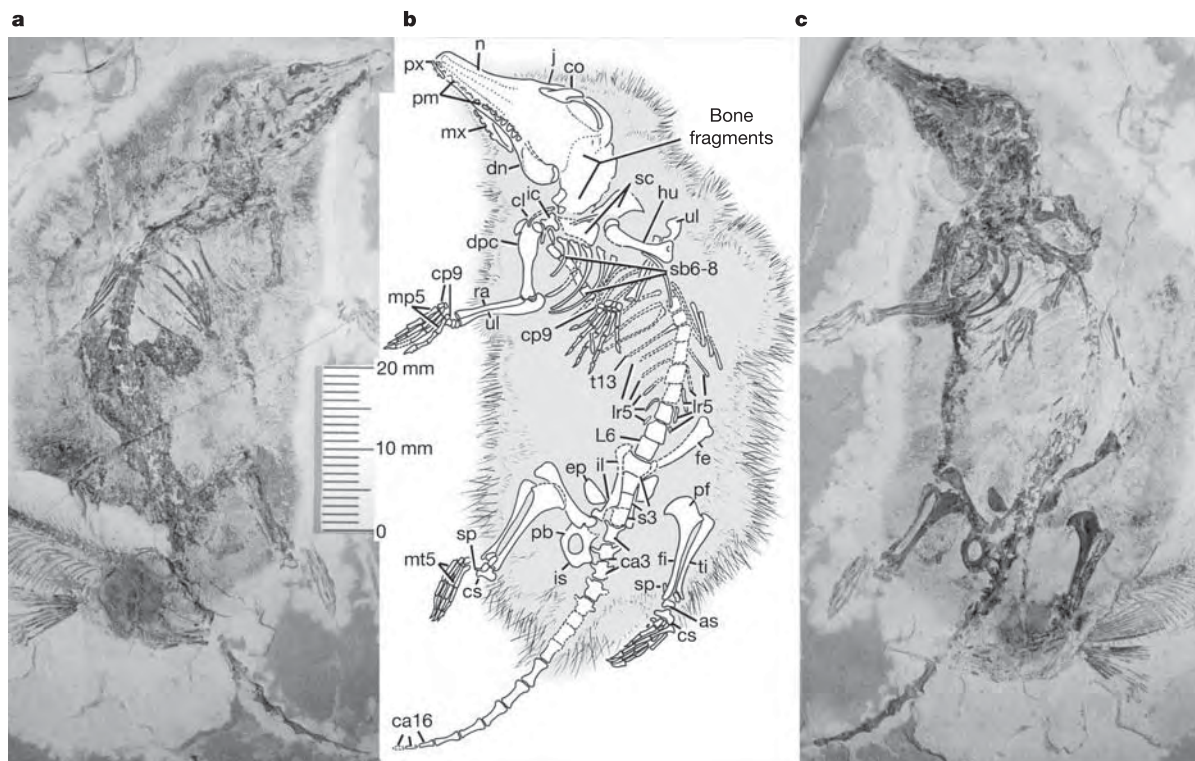


Figure 1 | *Akidolestes cifellii*. **a, c**, Counterpart (**a**) and main part (**c**) of the holotype (Nanjing Institute of Geology and Palaeontology, Academia Sinica, NIGPAS139381A, B). **b**, Skeletal features and fur outline of NIGPAS139381A. Abbreviations: as, astragalus; ca3, caudal vertebrae 1 through 3; ca16, caudal vertebrae 14 through 16; cl, clavicle; cp9, carpals 1 through 9; co, coronoid process of the dentary; cs, calcaneus; dn, dentary; dpc, deltopectoral crest (humerus); ep, epipubis; fe, femur; fi, fibula; hu, humerus; ic, interclavicle; il, ilium; is, ischium; j, jugal; L6, lumbar vertebrae

1 through 6; lr5, lumbar ribs 1 through 5; mp5, metacarpals 1 through 5; mt5, metatarsals 1 through 5; mx, broken and separated maxilla with upper molars; n, nasal; pb, pubis; pf, parafoveolar process of fibula; pm, lower premaxilla; px, broken and separated premaxilla with incisors; ra, radius; s3, sacral vertebrae 1 through 3; sc, scapula; sp, extratarsal 'poison' spur including os calcis and cornu calcis; sb6–8, sternbrae 6 through 8 (including xiphoid); ti, tibia; t13, the 13th thoracic rib (left); ul, ulna.

¹Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing 210008, China. ²Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213, USA. ³Department of Earth Sciences, Nanjing University, Nanjing 200017, China.

Holotype. Nanjing Institute of Geology and Palaeontology, Nanjing, China (NIGPAS) 139381A, B (Fig. 1), a skeleton with partial skull and dentition preserved in part and counterpart.

Etymology. *Akidolestes*: *akido-* (Greek) for point, for the pointed rostrum of this new mammal; *-lestes* (Greek), for thief, a common suffix for the name of fossil mammals; *cifellii*, in honour of Richard L. Cifelli, for his pioneering studies of symmetrodont mammals.

Locality, age and associated fauna. Yixian lacustrine beds at the Dawangzhangzi Locality, Lingyuan, Liaoning, China. The locality is correlated with other localities in Liaoning dated to be 124.6 Myr of the Barremian stage of the Lower Cretaceous^{1,2}, although there is no universal agreement on correlating the Yixian Formation to the European marine stages³. Other mammals of this formation include eutriconodontans^{14,15}, multituberculates⁴, symmetrodonts^{5–7}, metatherians¹⁶ and eutherians¹⁷.

Diagnosis. Symmetrodont with dentition of I4.C1.P5(?).M5(?)/i4.c1.p5.m6, with successively more acute angles of cusps from posterior premolars to posterior molars, in which cusp angles are less than 50° (Fig. 2). Molars with acute-triangulation of cusps and other features are typical of spalacotheroids that include zhangheotheriids^{5–7} and spalacotheriids^{8–11}; differs from *Zhangheotherium*^{5,6} and *Maothierium*⁷ of the Yixian Formation in having higher protocristid on molars, longer (larger) posterior premolars than anterior molars, and more premolars; from *Symmetrolestes*⁹ in having more molars; from older *Spalacotherium*¹⁰ and younger *Spalacolestes*⁸ and *Heinshanlestes*¹¹ in having a gracile coronoid process of the mandible (although similar to zhangheotheriids in this feature). *Akidolestes* is more primitive, in retaining distinctive cusps on the ultimate lower molar with symmetrical crown, than the geologically younger and derived spalacolestines, which lack cusp separation on

an asymmetrical ultimate lower molar^{8,9,11}. *Akidolestes* is also distinguishable from all other Mesozoic mammaliaforms, including the paraphyletic “obtuse-angled symmetrodonts”^{9,18}, in a combination of primitive and derived features to be described below (see also Supplementary Information).

Description. The mandible of *Akidolestes cifellii* is nearly identical to those of *Zhangheotherium*^{5,6} and *Maothierium*⁷ in having an elongate and gracile coronoid process and a mediolaterally compressed dentary condyle. However, the anterior portion of the mandible is more gracile, corresponding to the anteriorly narrow upper jaws and rostrum (Fig. 1), and differing from the broader rostrum of these zhangheotheriids⁷. The lower molar has a lower, continuous prevallid shearing surface between the protoconid and the paraconid and a higher, continuous postvallid shearing surface between the protoconid and metaconid. This is more derived than zhangheotheriids with the interrupted prevallid and postvallid surfaces, but similar to spalacotheriids^{8–11}. *A. cifellii* differs from zhangheotheriids but is very similar to spalacotheriids^{8–11} in having large posterior premolars that are longer than molars (Fig. 2). *Akidolestes* is unequivocally placed within the family Spalacotheriidae by dental characteristics (Fig. 3b, node 7).

The shoulder girdle and forelimb are similar to those of zhangheotheriids^{5–7}. However, *Akidolestes* differs from zhangheotheriids but is similar to monotremes in many features in the posterior part of the skeleton^{12,13} (Figs 1, 2, 4). Of the six lumbar vertebrae, five have unfused ribs (Fig. 2d, e), similar to the condition of monotremes, the eutriconodont *Repenomamus*¹⁵, *Fruitafossor*¹⁹ and many pre-mammaliaform cynodonts^{20,21}. The presence of mobile lumbar ribs differs conspicuously from the absence of these ribs in the closely related zhangheotheriids^{5–7} and some Mesozoic mammals, or their fusion to the lumbar centra in other Mesozoic mammals^{4,14,16,17,22–24}.

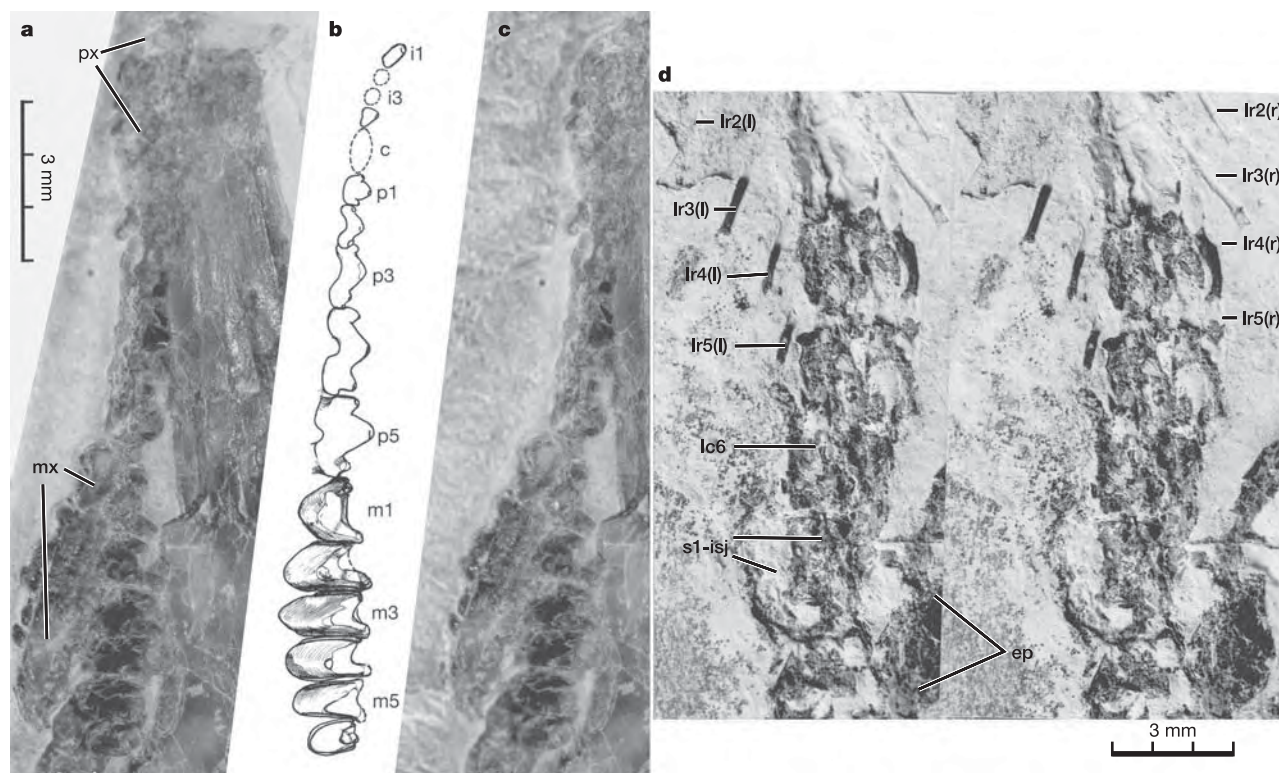


Figure 2 | Dentition of *Akidolestes cifellii*. **a, c**, Stereo photograph of the left lower teeth (**a**) and incomplete upper teeth (**c**) of NIGPAS139381A. **b**, Composite reconstruction of the lower teeth on the main part (NIGPAS139381A) and impression on the counterpart (NIGPAS139381B). **d**, Mobile lumbar ribs (stereo photograph of NIGPAS139381A; preserved on NIGPAS139381B but not illustrated). Abbreviations: ep, plate-like epipubis;

lc6, lumbar centrum 6; lr2–5(l), left lumbar ribs 2–5; lr2–5(r), right lumbar ribs 2–5; mx, broken maxilla with five preserved molars; px, broken premaxilla with upper incisors and incisor alveoli (maxilla and premaxilla are separated from cranium by the lower jaw); s1–isj, sacral vertebra 1 and ilio-sacral joint (outline on NIGPAS139381A, broken bone on NIGPAS139381B).

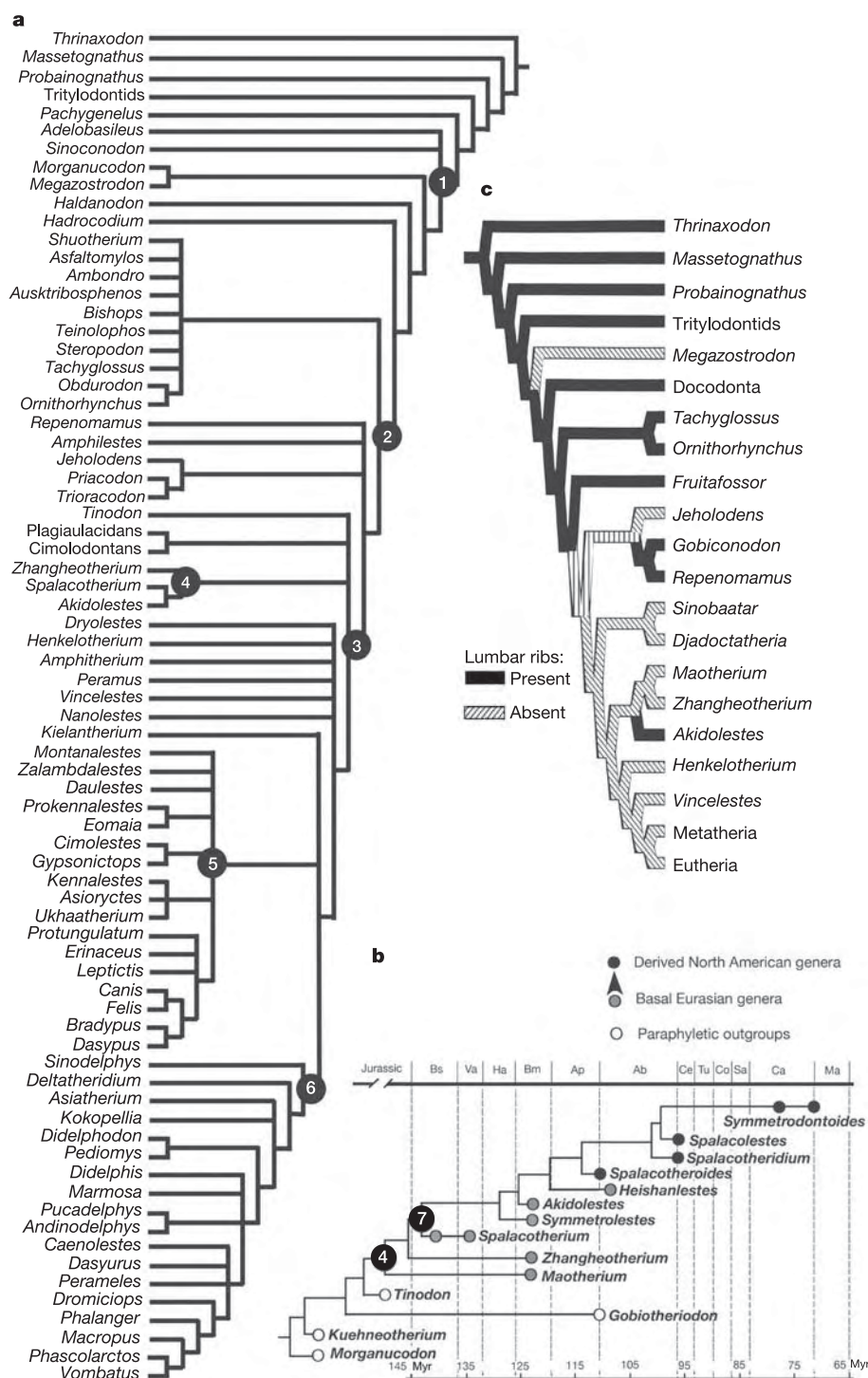


Figure 3 | Phylogenetic relationships of *Akidolestes cifellii*. **a**, Relationship of *A. cifellii* to major mammaliaform clades. **b**, Relationship of *A. cifellii* to other spalacotheroids (= basal trechnotherians). **c**, Hypothesis on homoplasy of lumbar ribs among mammaliaforms in which the intact lumbar region is preserved (tree simplified from **a** with data from ref. 19). Black branches, lumbar ribs present; hatched branches, lumbar ribs absent. The mammaliaform phylogeny is based on the strict consensus of 200 equally parsimonious and shortest trees (tree length 1,819, consistency index 0.426, retention index 0.794) from a PAUP analysis (version 4.0b; 1,000 runs of heuristic search with unordered multistate characters) of 413 morphological characters (from refs 16–19) that can be scored for the 74 comparative clades (including 5 pre-mammaliaform cynodont as outgroups and 17 extant mammal genera). Placement of *Akidolestes* within

spalacotheroids (node 4) and spalacotheriids (node 7) is based on a single shortest tree from 28 dental and mandibular characters of 10 spalacotheroid genera^{8,9} (tree length 47, consistency index 0.702, retention index 0.821, PAUP branch and bound search). Tree nodes: 1, Mammaliaformes; 2, Mammalia; 3, Theriiformes; 4, Spalacotheroidea; 5, Eutheria; 6, Metatheria; 7, Spalacotheriidae. Temporal distribution of spalacotheroids follows refs 4–11; the pattern of the geodispersal of spalacotheroids is consistent with Eurasia–North America dispersal patterns of all major groups that are common in Eurasian and North American Cretaceous faunas^{9,28,29}. Cretaceous stages shown in **b**: Ab, Albian; Ap, Aptian; Bm, Barremian; Bs, Berriasian; Ca, Campanian; Ce, Cenomanian; Co, Coniacian; Ha, Hauterivian; Ma, Maastrichtian; Sa, Santonian; Tu, Turonian; Va, Valanginian.

On the pelvis (Fig. 4), the epipubic bone is a broad plate, similar to that of *Ornithorhynchus* but different from the narrow epipubes of *Tachyglossus* and other Mesozoic mammals^{4–7,16,17}. The pubis has a prominent tubercle for the psoas minor muscle (Fig. 4a), a feature otherwise present only in living monotremes¹³ but absent in living therians and all other Mesozoic mammals for which the pelvis is known. In the head, neck and trochanteric area of the femur, *Akidolestes* is most similar to morganucodontans²⁴, eutriconodontans^{14,15} and monotremes (although to a smaller extent for the latter), but different from the closely related zhangheotheriids (Fig. 4), multituberculates and cladotherians^{4,16,17}. A striking feature of *Akidolestes* is a hypertrophied parafibular process of the fibula. The parafibula is a homoplastic feature, formed from a small ossification independent of the fibular diaphysis in some marsupials; it shows

variation in multituberculates^{22,23} but the parafibula is hypertrophied and fused completely to the fibula in living monotremes.

Akidolestes is similar to monotremes in the hindlimb. During the entire propulsive phase of locomotion in extant monotremes, the femur is horizontal and abducted, with a flexed knee joint^{12,13}. This sprawling posture is correlated with the hypertrophied parafibular process, which is so large so that it constrains the knee joint to be permanently flexed in abducted position^{12,13} (Fig. 4j). The sprawling posture is also correlated with a short femoral neck and a curved tibia with a distal malleolus for an asymmetrical upper ankle joint (Fig. 4d, dtm). We postulate that *Akidolestes* had a sprawling hindlimb posture from the similar osteological correlates of the sprawling posture of monotremes (Fig. 4i, j).

By contrast, the hindlimb posture of *Zhangheotherium* is more

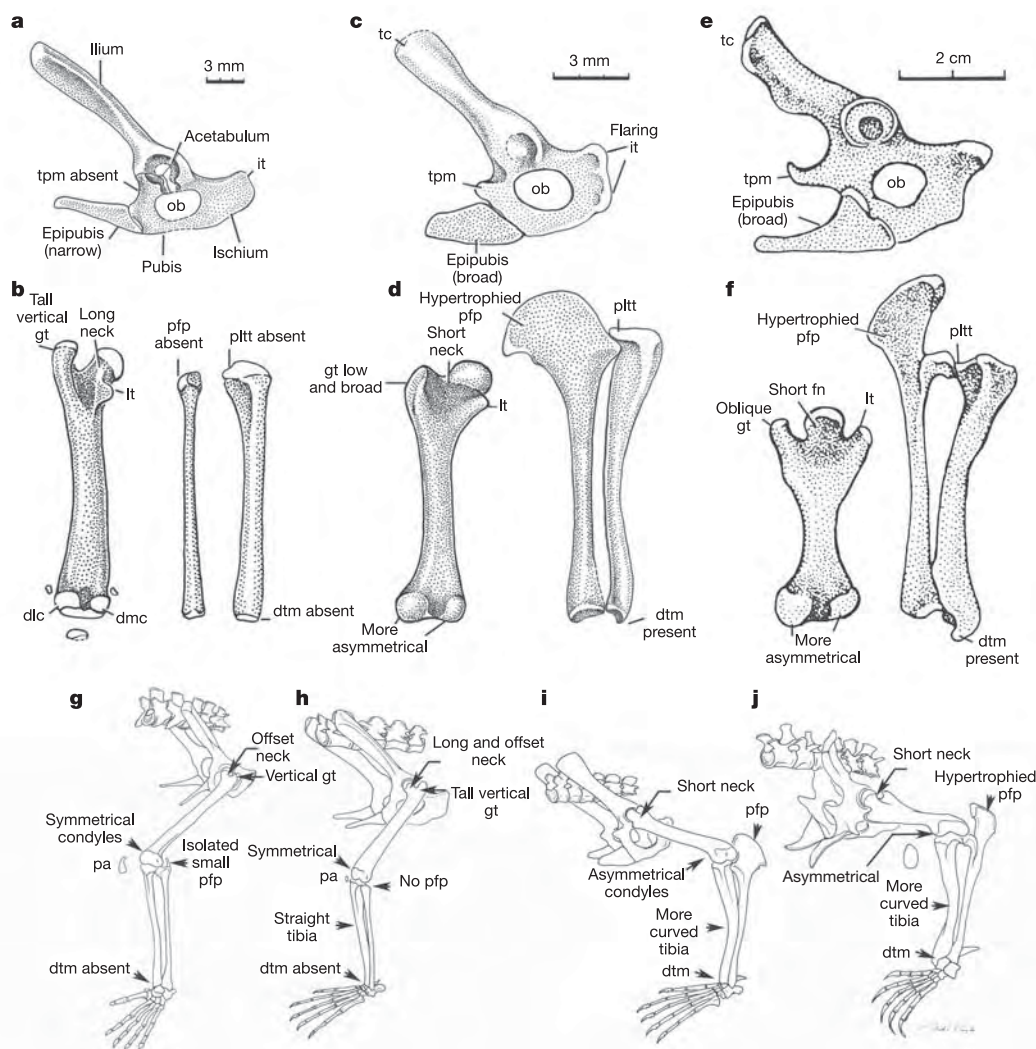


Figure 4 | Comparison of pelvic and hindlimb features of the spalacotheroids *Akidolestes* and *Zhangheotherium*, and hindlimb posture of *Didelphis* and *Ornithorhynchus*. **a, b, *Zhangheotherium*:** left pelvis (ventrolateral view) (**a**); left femur (posterior view), right fibula and tibia (lateral to anterolateral view) (**b**). **c, d, *Akidolestes*:** pelvis (ventrolateral view) (**c**); femur (posterior view), fibula and tibia (both in lateral view) (**d**). **e, f, *Ornithorhynchus*:** pelvis (ventrolateral view) (**e**); femur (posterior view), fibula and tibia (lateral view) (**f**). **g, Near-parasagittal hindlimb posture of the opossum *Didelphis*** (anterolateral view of the pelvis and hindlimb; arrows indicate the key characters for a more erect posture). **h, Hindlimb posture of *Zhangheotherium*** (more similar to opossum than to monotremes). **i, Hindlimb posture of *Akidolestes*** (more similar to monotremes than to opossum). **j, *Ornithorhynchus*:** anterolateral view of the

pelvis and hindlimb; arrows indicate the key characters for sprawling posture). Abbreviations: dlc, distal lateral condyle; dmc, distal medial condyle (of femur); dtm, distal tibial malleolus; ep, epipubis; neck, femoral neck (distinctive and angled in *Zhangheotherium*; short and indistinct in *Akidolestes*); gt, greater trochanter (high and vertical in *Zhangheotherium*; triangular and broad in *Akidolestes*); it, ischial tuberosity; lt, lesser trochanter; ob, obturator foramen; pfp, parafibular process (hypertrophied, fused in *Akidolestes* and *Ornithorhynchus*; small and isolated in *Didelphis*; absent in *Zhangheotherium*); pa, patella (relocated in illustration to show the distal femur); pltt, proximolateral tuberosity of tibia (large in *Akidolestes*); tpm, tubercle for M. psoas minor (on pubis); tc, tuber coxae (of ilium). For comparison of these pelvic and hindlimb features see Supplementary Information.

similar to that of *Didelphis* (Fig. 4g, h), on the basis of many osteological correlates for a more erect or parasagittal posture (Supplementary Information). The femur has a high and vertical greater trochanter, and a more distinctive neck, offset from the shaft (Fig. 4b, g). The distal femoral condyles are nearly equal; the fibula lacks the parafibular process; the tibia is straight. The long bones in zhangheotheriids^{5–7} would be oriented as in the extant *Didelphis* and derived cladotherians.

The phalangeal length ratios in each digit and profile of the terminal phalanx of *Akidolestes* differ from those of fossorial and semi-aquatic mammals so it would be unlikely to have had a fully fossorial adaptation¹⁹, or a fossorial and semi-aquatic adaptation²⁵. It lacks the phalangeal characteristics of scansorial mammals^{16,17}. On the basis of the structure of manus and pes, *Akidolestes* was most probably a generalized terrestrial mammal, like zhangheotheriids^{5–7} and morganucodontans²⁴.

The complete fossil of *Akidolestes* made it possible to evaluate these 'exceptional' features in the context of global parsimony (Fig. 3c). Our analyses of all features of *Akidolestes* have unequivocally placed it in the spalacotheroid clade within the trechnotherian group (Fig. 3a, b). Although 'unusual' for all theriiform mammals that are close relatives to *Akidolestes*, the lumbar ribs (Fig. 4) are clearly atavistic reversals to the primitive condition of the successively more distant groups of some (but not all) eutriconodontans, monotremes and nonmammalian cynodonts; the hypertrophied parafibula is convergent to those of distantly related monotremes. The mobile lumbar ribs are plesiomorphies of nonmammalian synapsids^{20,21}. Presence of these lumbar ribs in *Akidolestes*, which is nested deeply inside successive ranks of clades that do not have lumbar ribs (Fig. 3c), can be proposed as a phylogenetically homoplastic and functionally convergent feature or as the result of evolutionary development.

In extant monotremes, the posterior thoracic and anterior lumbar ribs provide attachment for many muscles of locomotory and respiratory functions¹³, including the following: the lumbar portion of the diaphragm for breathing; the psoas minor muscle inserting on the psoas minor tubercle for flexing the lumbar and pelvis; the psoas major muscle inserting on the lesser trochanter of the femur for rotating the femur; the quadratus lumborum muscle for flexing the lumbar and pelvic region; and the longissimus dorsi and iliocostalis lumborum muscles for extension of the lumbar and pelvic region¹³.

The presence of long lumbar ribs, a large psoas minor tubercle and the expanded anterior end of the ilium in *Akidolestes* indicate that the flexor and extensor muscles of the lumbar and pelvic region are well developed in this spalacotheriid, as in the extant monotremes, and more so than in *Zhangheotherium*, which is more similar to the marsupial *Didelphis* (Fig. 4, and Supplementary Information). The hypertrophied parafibula in *Akidolestes* would provide an expanded origination for several enlarged muscles for flexing the upper ankle joint and pedal digits, as in monotremes¹³. Given these many similarities, we infer that *Akidolestes* had a strong capacity for flexion and extension of the lumbar–pelvic region of the skeleton, for rotation of the femur, and for strong flexion of the pes, in convergence to the locomotory function of modern monotremes. The presence of the epipubic bone is correlated with the cross-couplet hypaxial muscle function in plesiomorphic locomotory pattern of basal mammals²⁶. However, it is difficult to interpret the homoplastic variation of the epipubic bone (large and broad versus gracile and small) in *Akidolestes* and other spalacotheroids in terms of convergent evolution of functionally adaptive features.

Within eutriconodontans, lumbar ribs are present in gobiconodontids but not in the related *Jeholodens*. Within spalacotheroids, these are present in *Akidolestes* but absent in zhangheotheriids. Outside the crown mammals, lumbar ribs are absent in morganucodontans²⁴ but variably present in many advanced cynodonts^{20,21}. It is possible that this rampant homoplasy of the lumbarosacral vertebral ribs is patterned by developmental genes that are deeply

conserved in widely separated mammalian taxa that lacked a recent common history²⁷. However, homoplastic development of the lumbar ribs is not mutually exclusive of the interpretation that these ribs and related features also have convergent function to extant monotremes.

Mammalian biogeography of Laurasia during the Early Cretaceous is characterized by iterative dispersals of major clades from their ancestral area of Asia to North America, where arrival of immigrant lineages is correlated with rapid turnover of the mammalian faunas. The phylogeny of spalacotheroids, including newly discovered taxa such as *Akidolestes*, suggests that basal spalacotheroid taxa are entirely Eurasian during the Berrasian–Barremian ages of the Cretaceous^{5–11,28} (Fig. 3b), and younger and more derived taxa are North American⁸. This geodispersal is consistent with palaeobiogeographical patterns of the main clades of Cretaceous mammalian faunas of Eurasia and North America, including multituberculates⁴, eutriconodontans²⁸, eutherians^{17,28,29} and metatherians^{16,28}. The concordant geodispersal patterns³⁰ of unrelated lineages suggests that during the Early Cretaceous (Fig. 3b), Asia was a source area for the origination and emigration of the main mammalian groups that became the major elements in North American faunas of the Late Cretaceous^{6,16,17,28,29}.

Received 21 April; accepted 24 August 2005.

- Ji, Q. et al. *Mesozoic Jehol Biota of Western Liaoning, China* (Geological Publishing House, Beijing, 2004).
- Zhou, Z.-H., Barrett, P. M. & Hilton, J. An exceptionally preserved Lower Cretaceous ecosystem. *Nature* **421**, 807–814 (2003).
- Chen, P.-J. et al. On the Jianshangou Beds of the Yixian Formation. *Science China D* **34**, 883–895 (2004).
- Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs—Origins, Evolution, and Structure* (Columbia Univ. Press, New York, 2004).
- Hu, Y.-M., Wang, Y.-Q., Luo, Z.-X. & Li, C.-K. A new symmetrodont mammal from China and its implications for mammalian evolution. *Nature* **390**, 137–142 (1997).
- Luo, Z.-X. & Ji, Q. New study on dental and skeletal features of the Cretaceous mammal *Zhangheotherium*. *J. Mamm. Evol.* **12**, 337–357 (2005).
- Rougier, G. W., Ji, Q. & Novacek, M. J. A new symmetrodont mammal from the Mesozoic of China. *Acta Geol. Sin.* **77**, 7–14 (2003).
- Cifelli, C. L. & Madsen, S. K. Spalacotheriid symmetrodonts (Mammalia) from the medial Cretaceous (Upper Albion or Lower Cenomanian) Mussentuchit local fauna, Cedar Mountain Formation, Utah, USA. *Geodiversitas* **21**, 167–214 (1999).
- Tsubamoto, T., Rougier, G. W., Isaji, S., Manabe, M. & Forasipai, A. M. New Early Cretaceous spalacotheriid 'symmetrodont' mammal from Japan. *Acta Palaeontol. Polonica* **49**, 329–346 (2004).
- Gill, P. A new symmetrodont from the Early Cretaceous of England. *J. Vert. Paleontol.* **24**, 748–752 (2004).
- Hu, Y.-M., Fox, R. C., Wang, Y.-Q. & Li, C.-K. A new spalacotheriid symmetrodont from the Early Cretaceous of Northeastern China. *Am. Mus. Novit.* **3475**, 1–20 (2005).
- Pridmore, P. A. Terrestrial locomotion in monotremes (Mammalia: Monotremata). *J. Zool.* **205**, 53–73 (1985).
- Gambaryan, P. P., Aristov, A. A., Dixon, J. M. & Zubtsova, G. Y. Peculiarities of the hind limb musculature in monotremes: an anatomical description and functional approach. *Russ. J. Theriol.* **1**, 1–36 (2002).
- Ji, Q., Luo, Z.-X. & Ji, S.-A. A Chinese triconodont mammal and mosaic evolution of the mammalian skeleton. *Nature* **398**, 326–330 (1999).
- Hu, Y.-M., Meng, J., Wang, Y.-Q. & Li, C.-K. Large Mesozoic mammals fed on young dinosaurs. *Nature* **433**, 149–153 (2005).
- Luo, Z.-X., Ji, Q., Wible, J. R. & Yuan, C.-X. An Early Cretaceous tribosphenic mammal and metatherian evolution. *Science* **302**, 1934–1940 (2003).
- Ji, Q. et al. The earliest-known eutherian mammal. *Nature* **416**, 816–822 (2002).
- Luo, Z.-X., Kielan-Jaworowska, Z. & Cifelli, R. L. In quest for a phylogeny of Mesozoic mammals. *Acta Palaeontol. Pol.* **47**, 1–78 (2002).
- Luo, Z.-X. & Wible, J. R. A new Late Jurassic digging mammal and early mammalian diversification. *Science* **308**, 103–107 (2005).
- Jenkins, F. A. Jr The postcranial skeleton of African cynodonts. *Peabody Mus. Nat. Hist. Bull.* **36**, 1–216 (1971).
- Sues, H.-D. *Advanced Mammal-like Reptiles from the Early Jurassic of Arizona* Thesis, Harvard Univ. (1983).
- Krause, D. W. & Jenkins, F. A. Jr The postcranial skeleton of North American multituberculates. *Bull. Mus. Comp. Zool.* **150**, 199–246 (1983).
- Kielan-Jaworowska, Z. & Gambaryan, P. P. Postcranial anatomy and habits of Asian multituberculate mammals. *Fossils Strata* **36**, 1–92 (1994).

24. Jenkins, F. A. Jr & Parrington, F. R. The postcranial skeletons of the Triassic mammals *Eozostrodon*, *Megazostrodon* and *Erythrotherium*. *Phil. Trans. R. Soc. Lond. B* **273**, 387–431 (1976).
25. Martin, T. Postcranial anatomy of *Haldanodon expectatus* (Mammalia, Docodonta) from the Late Jurassic (Kimmeridgian) of Portugal and its bearing for mammalian evolution. *Zool. J. Linn. Soc.* **145**, 219–248 (2005).
26. Reilly, S. M. & White, T. D. Hypaxial motor patterns and the function of epipubic bones in primitive mammals. *Science* **299**, 400–402 (2003).
27. Wellik, D. M. & Capecchi, M. R. *Hox10* and *Hox11* genes are required to globally pattern the mammalian skeleton. *Science* **301**, 363–367 (2003).
28. Averianov, A. O. & Archibald, J. D. Mammals from the Upper Cretaceous Aitym Formation, Kyzylkum Desert, Uzbekistan. *Cretaceous Res.* **24**, 171–191 (2003).
29. Clemens, W. A. Patterns of mammalian evolution across the Cretaceous–Tertiary boundary. *Mitteilungen Mus. für Naturk. Berlin Zool. Reihe* **77**, 175–191 (2001).
30. Lieberman, B. S. Paleobiogeography: the relevance of fossils to biogeography. *Annu. Rev. Ecol. Evol. Syst.* **34**, 51–69 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P.-J. Chen for generously providing this fossil for our study; A. R. Tabrum for his skilful preparation of the fossil; H.-C. Zhang, J.-G. Sha, X.-N. Yang and Y.-K. Shi for their help with this research; R. L. Cifelli, Q. Ji, Z. Kielan-Jaworowska, T. Martin and J. R. Wible for sharing ideas on early mammals; K. C. Beard, M. R. Dawson and T. Martin for discussion on mammalian evolutionary biogeography; M. R. Dawson and J. R. Wible for improving the manuscript; and M. A. Klingler for assistance with graphics. Support was provided by the National Natural Science Foundation of China, the National Science Foundation (USA) and the National Geographic Society and Carnegie Museum of Natural History (Z.-X.L.). G.L. acknowledges the Nanjing Institute of Geology and Palaeontology and the Institute's State Key Laboratory of Palaeobiology and Stratigraphy, and funding from the Ministry of Science and Technology of China ('973' project funding to C.-S. Wang).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Z.-X.L. (luoz@carnegiemnh.org).

Outbred embryos rescue inbred half-siblings in mixed-paternity broods of live-bearing females

Jeanne A. Zeh¹ & David W. Zeh¹

Females commonly mate with more than one male¹, and polyandry has been shown to increase reproductive success in many species^{2–4}. Insemination by multiple males shifts the arena for sexual selection from the external environment to the female reproductive tract, where sperm competition or female choice of sperm could bias fertilization against sperm from genetically inferior⁵ or genetically incompatible males^{6,7}. Evidence that polyandry can be a strategy for avoiding incompatibility comes from studies showing that inbreeding cost is reduced in some egg-laying species by postcopulatory mechanisms that favour fertilization by sperm from unrelated males^{8–10}. In viviparous (live-bearing) species, inbreeding not only reduces offspring genetic quality but might also disrupt feto-maternal interactions that are crucial for normal embryonic development^{11–13}. Here we show that polyandry in viviparous pseudoscorpions reduces inbreeding cost not through paternity-biasing mechanisms favouring outbred offspring, but rather because outbred embryos exert a rescuing effect on inbred half-siblings in mixed-paternity broods. The benefits of polyandry may thus be more complex for live-bearing females than for females that lay eggs.

Consanguineous mating is a significant source of fitness depression, and inbred offspring are more likely to be homozygous either for deleterious, recessive alleles or at loci with heterozygote advantage¹⁴. In live-bearing species, inbreeding can also disrupt the complex sequence of two-way, immunological interactions between fetal and maternal tissues¹⁵. Unfortunately, investigating inbreeding effects on fetal loss is hindered in most viviparous animals by the intrusive methods required to detect early stage, spontaneous abortion. Unlike most terrestrial arthropods, pseudoscorpions are viviparous. Embryos develop in an external, translucent brood sac and draw nutritive fluid from the mother's reproductive tract¹⁶. Previous research exploited this 'external-womb' form of viviparity to establish that polyandry in the pseudoscorpion, *Cordylodermes scorpoides*, significantly enhances female lifetime reproductive success by reducing the rate of spontaneous abortion of entire broods¹⁷. Here we investigate the effect of inbreeding on abortion rate and reproductive success in *C. scorpoides*, and assess whether females reduce this cost by mating with both a related and an unrelated male.

Virgin females were randomly assigned to one of five treatments, four involving a single mating with each of two males: two non-siblings of the female (NN), two full-siblings of the female (SS), a non-sibling first male and a full-sibling second male (NS), or vice versa (SN). Unlike the SS treatment, male pairs in NN replications were non-brothers, the usual situation for polyandrous females in nature¹⁸. For a conservative comparison, a fifth treatment included females that were mated twice to a single non-sibling (N; see Methods). Each mated female was monitored until she gave birth to a first brood of nymphs, spontaneously aborted her first brood, or failed to produce a brood in 45 d. All nymphs were counted at birth,

and DNA profiling¹⁹ was used to assign paternity for NS and SN nymphs. A second experiment assessed the effect of relatedness on number of sperm males allocated to spermatophores. Once-mated males were paired with a full-sibling (S female) or a non-relative (N female; $n = 12$ per treatment), the sperm packet was collected, and sperm were counted¹⁷.

A randomized block analysis of covariance (ANCOVA) showed a significant effect of mating treatment ($F_{4,187} = 8.45$, $P < 0.001$) and female body size ($F_{1,192} = 7.55$, $P = 0.007$) on reproductive success, that is, number of nymphs born. Comparison of treatment means (Fig. 1) showed that female reproductive success sorted into three levels, $SS < NS = SN = N < NN$, a grouping that was highly significant (contrast analysis²⁰, $F_{1,187} = 22.12$, $P < 0.001$), and accounted for 65% of the mating treatment effect. Underlying this pattern was a significant treatment effect on the frequency of spontaneous brood abortion (Fisher exact test, $P = 0.029$), with SS females suffering the highest abortion rate (40%; see Fig. 1). Mating treatment also significantly influenced the number of nymphs born to females that successfully carried broods to term ($F_{4,144} = 7.10$, $P < 0.001$); the $SS < NS = SN = N < NN$ pattern was again significant ($F_{1,144} = 13.93$, $P < 0.001$), accounting for 49% of the treatment effect. Thus, mating with a non-relative and a brother diminished cost of inbreeding by both reducing the risk of spontaneous brood abortion and increasing embryonic survival in successful broods. In the second experiment, relatedness had no effect on number of sperm allocated to spermatophores (S-female,

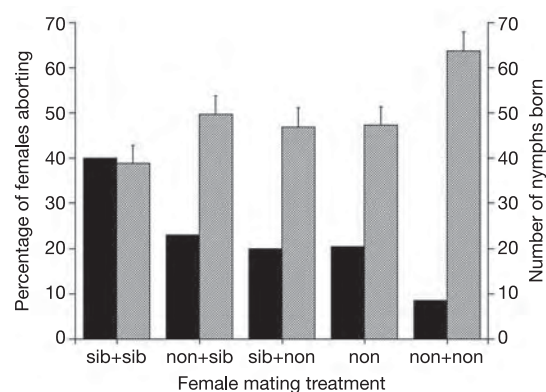


Figure 1 | Effect of mating treatment on brood production. Rates of whole-brood spontaneous abortion (filled bars) and number of nymphs (mean $\pm$ s.e.m.) born to females carrying broods to term (hatched bars) for treatments in which females were mated to two full-siblings (sib + sib), to a non-relative followed by a full-sibling (non + sib), to a full-sibling followed by a non-relative (sib + non), to a single non-relative (non), or to two non-relatives (non + non).

¹Department of Biology and Program in Ecology, Evolution and Conservation Biology, University of Nevada, Reno, Nevada 89557, USA.

1,674 $\pm$ 147; N-female, 1,637 $\pm$ 151 (mean $\pm$ s.e.m.); $t_{22} = 0.18$, $P = 0.86$).

The SS < NS = SN = N < NN pattern of female reproductive success is consistent with the hypothesis that polyandry enables *C. scorpoides* females to bias paternity against sibling males. However, paternity assignment for NS and SN nymphs showed that this was not the case. Regardless of mating order, females gave birth to more inbred than outbred offspring. The proportion of NS and SN inbred nymphs was 0.66 $\pm$ 0.17 and 0.60 $\pm$ 0.24 (mean $\pm$ 95% confidence interval), respectively, and did not differ significantly between treatments ($F_{1,31} = 0.19$, $P = 0.668$). The pooled mean of 63% ($\pm$ 14%) suggests a sperm-competition advantage for sibling males. Because of higher mortality of inbred embryos, this at-birth value of 63% underestimates the actual fertilization success of sibling males. After correcting for embryonic survival differences (Supplementary Information), sibling-male fertilization success was more than twice that of non-relative males (69% $\pm$ 14% versus 31% $\pm$ 14%, $P = 0.011$). Although the mechanism underlying the sperm-competition advantage is unknown, genetic similarity could feasibly lower the risk for sibling-male sperm of being targeted by female immune defences.

This sibling-male, sperm-competition advantage can have deleterious, postfertilization consequences for female reproductive success. In 12 cases, females produced only inbred nymphs (Fig. 2). These females gave birth to significantly fewer nymphs (35.25 $\pm$ 4.96; mean $\pm$ s.e.m.) than females whose broods included at least one outbred nymph (54.90 $\pm$ 3.84; $t_{30} = 3.13$, $P = 0.004$). The high rate of spontaneous brood abortion and small brood size associated with inbreeding might result from males donating less sperm and/or accessory gland proteins (Acps) to related females²¹; however, the evidence available does not support this hypothesis. Males did not discriminate against sisters, either in propensity to produce a spermatophore or in number of sperm allocated to sperm packets. Adjustment of Acp quality and/or quantity independently of sperm number would therefore be required for Acp-mediated treatment effects on female reproductive success. Moreover, although seminal products stimulate ovulation in some live-bearing species,

mating treatment had no effect on whether *C. scorpoides* females produced a brood (Fisher exact test, $P = 0.2150$). Finally, with a mean ejaculate mass estimated at 1 μ g and a mean brood mass of 8,800 μ g just before birth, the material contribution made by males seems negligible.

Our findings show that polyandry can reduce the high rate of spontaneous brood abortion and the low reproductive success associated with inbreeding, if inbred embryos develop with outbred half-siblings. This rescuing effect can be quantified by considering reproductive success from a sibling-male perspective (Supplementary Information). A male that mates with his sister sires nearly twice as many inbred offspring that survive to birth if she also mates with an unrelated male (NS and SN treatments, 21.52 $\pm$ 2.28; mean $\pm$ s.e.m.), as compared with a second brother (SS treatment, 12.70 $\pm$ 2.55; Mann–Whitney test, $U = 1097.5$, $P = 0.021$). Indeed, despite losing about a third of fertilizations to the unrelated male, the reproductive success of the single, sibling male in NS and SN replications is statistically indistinguishable from the combined reproductive success (23.12 $\pm$ 3.86) of the two brothers in SS replications (Mann–Whitney test, $U = 1478$, $P = 0.990$), owing to the reduced rate of spontaneous brood abortion.

How might rescuing of inbred embryos occur? With inbreeding, homozygosity for a deleterious, recessive allele would result in intrinsically weak offspring that are unable to sequester adequate maternal resources. By their more vigorous activity, outbred embryos might draw sufficient nutrients into the communal brood sac to ensure development of the brood to birth. Alternatively, genetic similarity of inbred embryos to their mother may blur the self/non-self distinction essential for innate immune responses. Innate immunity, common to invertebrates and vertebrates²², has an important role in fetal loss, with perturbations of immune responses triggering spontaneous abortion²³. By establishing a non-self presence in mixed inbred/outbred broods and activating the normal cascade of feto-maternal interactions, outbred embryos might compensate for attenuation of interactions between mother and inbred concepti. Notably, both these hypotheses account for the variation in inbreeding cost apparent between sibling pairs in Figs 1 and 2. At any locus, whether it influences embryonic viability or is involved in recognition, mendelian genetics dictate that full-siblings may share one, both or no alleles. The fitness consequences of inbreeding will thus vary, depending on the multilocus genotypes of the particular brother and sister involved.

Our results are consistent with the view that parent–offspring conflict over maternal resource allocation drives the evolution of feto-maternal interactions, with resistance by the mother to manipulation of her reproductive physiology by the embryo creating tension that is crucial for normal embryonic development²⁴. Relaxation of this tension, through increased homozygosity or genetic similarity, could explain both the high rate of spontaneous inbred-brood abortion and the rescuing effect of outbred, half-siblings detected in our study.

So far, evidence that polyandry reduces inbreeding costs has been mixed, leading some to question the importance of genetic incompatibility avoidance in the evolution of polyandry²⁵. However, most counterevidence comes from egg-laying species, with females limited to pre-fertilization mechanisms for inbreeding avoidance. We have shown in a viviparous species that mixed paternity rather than paternity biasing reduces inbreeding cost. If mixed-paternity rescuing effects occur in other live-bearing species, reproductive mode may well be an important factor influencing not only the potential sources of incompatibility between paternal and maternal genomes⁶, but also the range of postcopulatory defences available to polyandrous females.

METHODS

Pseudoscorpions. Pseudoscorpions were laboratory-reared F₂ and F₃ descendants of 82 field-inseminated *C. scorpoides* females collected in Panama; nymphs

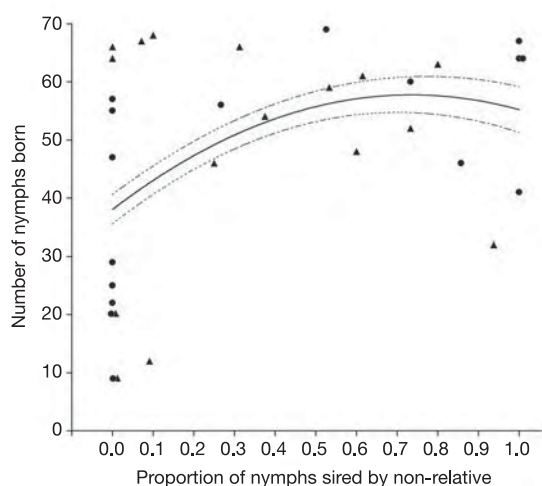


Figure 2 | Number of nymphs born versus proportion outbred in successful NS and SN broods. Cubic spline²⁷, standard and Gentleman–Givens regressions²⁶ all yielded a significant quadratic relationship between nymph number and proportion sired by the non-relative male in non+sib (triangles) and sib + non (circles) treatments. The unbroken line indicates the best-fit cubic spline solution. Broken lines indicate ± 1 s.e.m., based on 1,000 bootstrap replicates. For standard, quadratic regression, $R^2 = 0.239$, $P = 0.019$. Neither linear nor quadratic regression was significant when exclusively inbred broods were excluded ($R^2 = 0.009$, $P = 0.688$ and $R^2 = 0.013$, $P = 0.897$, respectively).

were reared individually to ensure virginity¹⁷. To generate full-sibling families, F₁ females were randomly mated to a single, unrelated male, and progeny were reared to adults. These F₂ individuals either were used in block 1 or were mated to provide F₃ individuals for block 2.

Polyandry and inbreeding experiment. Within blocks, a completely randomized mating design was used to maximize sample size. Before the experiment, each male was mated once to a non-experimental female. Each mating pair was placed in a 28-mm-diameter arena, interactions were videotaped for 30 min¹⁷, and pseudoscorpions were then returned to their vials. After 48 h, females were presented with the same male (N treatment) or a different male (NN, NS, SN and SS treatments). Only females that unambiguously accepted a sperm packet in both matings were retained for subsequent monitoring ($n = 41$ for the N treatment; $n = 39$ for NN; $n = 39$ for NS; $n = 36$ for SN, and $n = 44$ for SS).

Monitoring female reproductive success. Females were maintained in transparent vials containing *Ficus* frass, were fed nine *Tribolium confusum* larvae per week, and were checked every 3–4 d, increasing to daily at late stages of gestation. During gestation, females remain in a silken nest constructed on the vial wall, enabling non-intrusive monitoring of brood development¹⁷. After a 14-d developmental period, nymphs are born simultaneously and remain in the nest for 2–3 d. Within 24 h of birth, nymphs were counted, and NS and SN nymphs were frozen for paternity testing. Females were assigned a reproductive success score of 0 if they aborted their entire first brood. Females not producing a brood sac in 45 d ($n = 11$) were excluded from the reproductive success analyses, because such females do not become gravid¹⁷.

Statistical analyses were done with SAS²⁶. Cephalothorax length, the trait most strongly correlated with female lifetime reproductive success¹⁷, was used to control for body size in the ANCOVA analysis of female reproductive success (see text). For the analysis of the complete data set, a non-parametric ANCOVA was done on rank-transformed reproductive success data and residuals were normally distributed (Shapiro-Wilk test, $W = 0.989$, $P = 0.157$). For the analysis excluding cases of spontaneous abortion, data transformation was not required ($W = 0.990$, $P = 0.403$).

Minisatellite paternity assignment. Amplification by PCR of alleles at the *cCscMS23* minisatellite locus (heterozygosity = 0.99)¹⁹ was used to assign paternity for 465 nymphs from 16 replications of each of the NS and SN treatments. PCR products from the mother, the two putative sires and 8–20 offspring were run on agarose gels stained with ethidium bromide to visualize alleles. Paternity was assigned on the basis of the presence of unique paternal alleles in offspring. Only replications in which the putative sires shared no alleles were scored (32 out of 34 replications tested).

Several measures ensured amplification of both alleles from heterozygous individuals (Supplementary Information). First, a strict consensus primer pair was designed from a large sample of *cCscMS23* alleles ($n = 22$). Second, the high primer-pair T_m enabled a robust, two-step PCR reaction. Finally, PCR was done with a mixture of DNA polymerases capable of amplifying high molecular mass alleles (to 22 kb) from femtomgram quantities of DNA.

Received 14 April; accepted 21 September 2005.

- Jennions, M. D. & Petrie, M. Why do females mate multiply? A review of the genetic benefits. *Biol. Rev. Camb. Phil. Soc.* **75**, 21–64 (2000).
- Hasselquist, D., Bensch, S. & von Schantz, T. Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Nature* **381**, 229–232 (1996).
- Tregenza, T. & Wedell, N. Polyandrous females avoid the costs of inbreeding. *Nature* **415**, 71–73 (2002).
- Stockley, P. Female multiple mating behaviour, early reproductive failure and litter size variation in mammals. *Proc. R. Soc. Lond. B* **270**, 271–278 (2003).
- Birkhead, T. R., Møller, A. P. & Sutherland, W. J. Why do females make it so difficult for males to fertilize their eggs? *J. Theor. Biol.* **161**, 51–60 (1993).
- Zeh, J. A. & Zeh, D. W. Reproductive mode and the genetic benefits of polyandry. *Anim. Behav.* **61**, 1051–1063 (2001).

- Tregenza, T. & Wedell, N. Genetic compatibility, mate choice and patterns of parentage: invited review. *Mol. Ecol.* **9**, 1013–1027 (2000).
- Olsson, M., Shine, R., Madsen, T., Gullberg, A. & Tegelström, H. Sperm selection by females. *Nature* **383**, 585 (1996).
- Foerster, K., Delhey, K., Johnsen, A., Lifjeld, J. T. & Kempenaers, B. Females increase offspring heterozygosity and fitness through extra-pair matings. *Nature* **425**, 714–717 (2003).
- Bretman, A., Wedell, N. & Tregenza, T. Molecular evidence of post-copulatory inbreeding avoidance in the field cricket *Gryllus bimaculatus*. *Proc. R. Soc. Lond. B* **271**, 159–164 (2004).
- Ober, C. *et al.* Variation in the HLA-G promoter region influences miscarriage rates. *Am. J. Human Genet.* **72**, 1425–1435 (2003).
- Choudhury, S. R. & Knapp, L. A. Human reproductive failure II: immunogenetic and interacting factors. *Human Reprod. Update* **7**, 135–160 (2001).
- Davies, C. J. *et al.* Major histocompatibility antigen expression on the bovine placenta: its relationship to abnormal pregnancies and retained placenta. *Anim. Reprod. Sci.* **82–83**, 267–280 (2004).
- Keller, L. F. & Waller, D. M. Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**, 230–241 (2002).
- Hill, J. A. in *Immunology of Human Reproduction* (eds Kurpisz, M. & Fernandez, N.) 401–424 (Bios Scientific, Oxford, UK, 1995).
- Weygoldt, P. *The Biology of Pseudoscorpions* (Harvard Univ. Press, Cambridge, MA, 1969).
- Newcomer, S. D., Zeh, J. A. & Zeh, D. W. Genetic benefits enhance the reproductive success of polyandrous females. *Proc. Natl Acad. Sci. USA* **96**, 10236–10241 (1999).
- Zeh, D. W., Zeh, J. A. & Bermingham, E. Polyandrous, sperm-storing females: carriers of male genotypes through episodes of adverse selection. *Proc. R. Soc. Lond. B* **264**, 119–125 (1997).
- Zeh, D. W., Zeh, J. A. & May, C. A. Charomid cloning vectors meet the pedipalpal chelae: single-locus minisatellite DNA probes for paternity assignment in the harlequin beetle-riding pseudoscorpion. *Mol. Ecol.* **3**, 517–522 (1994).
- Rosenthal, R. & Rosnow, R. L. *Contrast Analysis. Focused Comparisons in the Analysis of Variance* (Cambridge Univ. Press, Cambridge, UK, 1985).
- Simmons, L. W. The evolution of polyandry: an examination of the genetic incompatibility and good-sperm hypotheses. *J. Evol. Biol.* **14**, 585–594 (2001).
- Hoffmann, J. A. The immune response of *Drosophila*. *Nature* **426**, 33–38 (2003).
- Caucheteux, S. M., Kanellopoulos-Langevin, C. & Ojcius, D. M. At the innate frontiers between mother and fetus: linking abortion with complement activation. *Immunity* **18**, 169–172 (2003).
- Haig, D. Genetic conflicts in human pregnancy. *Q. Rev. Biol.* **68**, 495–532 (1993).
- Jennions, M. D., Hunt, J., Graham, R. & Brooks, R. No evidence for inbreeding avoidance through postcopulatory mechanisms in the black field cricket, *Teleogryllus commodus*. *Evolution* **58**, 2472–2477 (2004).
- SAS Institute, *The SAS System for Windows, Release 8.02* (SAS Institute, Cary, 2001).
- Schluter, D. & Nychka, D. Exploring fitness surfaces. *Am. Nat.* **143**, 597–616 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Bonilla and S. Trimmer for assistance; K. Panter for discussion; the Smithsonian Tropical Research Institute for logistical support; and La Autoridad Nacional del Ambiente for permission to collect in Panama. This research was supported by grants from the US National Science Foundation and the National Geographic Society.

Author Contributions J.A.Z. designed the study, carried out the mating and rearing experiments, and took the primary role in writing the paper. D.W.Z. performed the molecular and statistical analyses.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A.Z. (jaz@unr.edu).

LETTERS

Specificity in Toll-like receptor signalling through distinct effector functions of TRAF3 and TRAF6

Hans Häcker^{1†}, Vanessa Redecke², Blagoy Blagoev⁴, Irina Kratchmarova⁴, Li-Chung Hsu¹, Gang G. Wang³, Mark P. Kamps³, Eyal Raz², Hermann Wagner⁵, Georg Häcker⁵, Matthias Mann⁴ & Michael Karin¹

Toll-like receptors (TLRs) are activated by pathogen-associated molecular patterns to induce innate immune responses and production of pro-inflammatory cytokines, interferons and anti-inflammatory cytokines¹. TLRs activate downstream effectors through adaptors that contain Toll/interleukin-1 receptor (TIR) domains², but the mechanisms accounting for diversification of TLR effector functions are unclear. To dissect biochemically TLR signalling, we established a system for isolating signalling complexes assembled by dimerized adaptors. Using MyD88 as a prototypical adaptor, we identified TNF receptor-associated factor 3 (TRAF3) as a new component of TIR signalling complexes that is recruited along with TRAF6. Using myeloid cells from TRAF3- and TRAF6-deficient mice, we show that TRAF3 is essential for the induction of type I interferons (IFN) and the anti-inflammatory cytokine interleukin-10 (IL-10), but is dispensable for expression of pro-inflammatory cytokines. In fact, TRAF3-deficient cells overproduce pro-inflammatory cytokines owing to defective IL-10 production. Despite their structural similarity, the functions of TRAF3 and TRAF6 are largely distinct. TRAF3 is also recruited to the adaptor TRIF (Toll/IL-1 receptor domain-containing adaptor-inducing IFN- β) and is required for marshalling the protein kinase TBK1 (also called NAK) into TIR signalling complexes, thereby explaining its unique role in activation of the IFN response.

TLRs function through four TIR domain adaptors that may act in pairs: MyD88 and TIRAP/MAL, and TRIF/TICAM-1 and TRAM¹. Whereas TLR4 uses all four, TLR9 signals via MyD88, and TLR3 mainly via TRIF^{3–5}. Gene targeting suggests that MyD88-dependent signalling requires TRAF6 (ref. 6), but how different TIR adaptors achieve diversification of effector functions is unclear. To answer this question we used a biochemical approach. To bypass receptor-induced dimerization, we fused MyD88 or the TRAF6 effector domain to subunit B of *Escherichia coli* DNA gyrase (GyrB; Fig. 1a and Supplementary Fig. 1a), which dimerizes upon binding coumermycin, a bivalent antibiotic⁷. Coumermycin treatment of cells expressing these constructs led to the activation of p38 mitogen-activated protein kinases (MAPK) (Fig. 1b), ERK1/2, JNK1/2 and I κ B kinases (IKK) (Supplementary Fig. 1 and data not shown). Activation-dependent degradation of IRAK1 was observed upon dimerization of MyD88–GyrB but not TRAF6–GyrB (Supplementary Fig. 1b), consistent with IRAK1 acting between MyD88 and TRAF6 (ref. 1). Dimerization of either MyD88–GyrB or TRAF6–GyrB induced TNF- α as efficiently as the TLR9 agonist CpG-DNA³ (Supplementary Fig. 1c). Hence, dimerization of adaptor proteins mimics TLR signalling.

After dimerization, MyD88–GyrB efficiently recruited TRAF6 (Fig. 1c). To isolate other MyD88-interacting proteins we established RAW cells expressing MyD88–GyrB fusion proteins containing an additional TAP (tandem affinity purification) tag (Fig. 1a)⁸. These cells and green fluorescent protein (GFP)–TAP-expressing control cells were stimulated with coumermycin for 6 min and lysates were prepared. Protein complexes were isolated by TAP, resolved by SDS–polyacrylamide gel electrophoresis (PAGE) and stained (Fig. 2a). Bands 1 and 2 are the bait proteins. Band 4 was constitutively associated with MyD88, whereas other, barely visible bands, including band 3, were dimerization-dependent. The identity of these bands was queried by mass spectrometry (MS)⁹. Band 4 contained IRAK4, an essential component of all TIR signalling complexes¹⁰. In band 3, TRAF6 was identified with 12 peptides. Surprisingly, band 3 also contained four TRAF3-derived specific peptides (Supplementary Fig. 2a). TRAF3 recruitment to MyD88–GyrB and interleukin-1 receptor (IL-1R) was confirmed by immunoprecipitation and was comparable to TRAF6 recruitment (Fig. 2b and Supplementary Fig. 2b). Moreover, activation of MyD88–GyrB transfectants with CpG-DNA or Pam₃Cys (TLR2)¹¹ induced association of endogenous TRAF3 with MyD88–GyrB (Fig. 2c and Supplementary Fig. 2c). Notably, lipopolysaccharide (LPS) did not induce detectable interaction between MyD88 and TRAF3 (Fig. 2c),

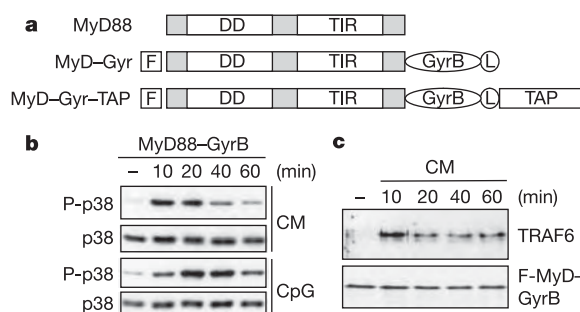


Figure 1 | Dimerization of MyD88 mimics TLR signalling. **a**, Constructs used. DD, death domain; TIR, TIR domain; GyrB, subunit B of *E. coli* DNA gyrase; L, IgG hinge region linker; TAP, TAP tag; F, Flag tag. **b**, MyD88–GyrB-expressing RAW264.7 cells were stimulated with coumermycin (CM) or CpG-DNA, and p38 phosphorylation was analysed by immunoblotting. **c**, Flag–MyD88–GyrB-expressing RAW264.7 cells were stimulated with coumermycin. Lysates were prepared, immunoprecipitated with anti-Flag antibodies and analysed by immunoblotting with antibodies to TRAF6 or MyD88.

¹Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology, School of Medicine, ²Department of Medicine, and ³Department of Pathology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. ⁴Center for Experimental Bioinformatics (CEBI), Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. ⁵Institute for Medical Microbiology, Immunology and Hygiene, Technische Universität München, Trogerstr. 9, D-81675 Munich, Germany. [†]Present address: Department of Infectious Diseases, St Jude Children's Research Hospital, 332 North Lauderdale Street, Memphis, Tennessee 38105, USA.

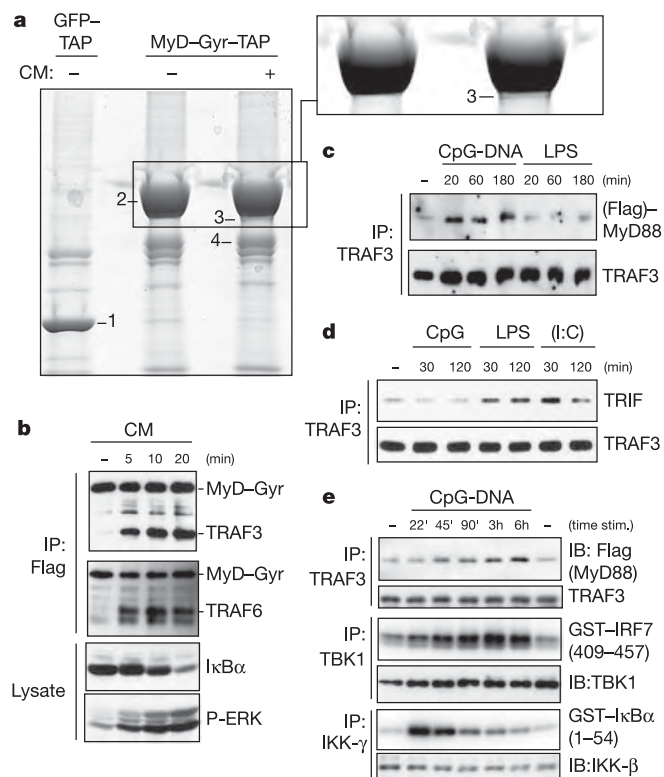


Figure 2 | TRAF3 is part of the TIR signalling complex. **a**, RAW264.7 cells expressing GFP-TAP or Flag-MyD88-GyrB-TAP were stimulated for 6 min with coumermycin as indicated. Cell lysates were prepared, TAP-purified, gel-separated and stained with Coomassie blue. Part of the gel is enlarged (right side). **b**, Flag-MyD88-GyrB stable transfectants were stimulated with coumermycin, and lysates were prepared and analysed directly or after immunoprecipitation (IP) with anti-Flag antibodies by immunoblotting (IB) as indicated. MyD88-GyrB was detected by TAP/protein A-dependent binding of antibodies used for immunoblotting. **c**, **d**, Flag-MyD88-GyrB stable transfectants were stimulated with CpG-DNA, LPS or poly(I:C). Lysates were prepared, immunoprecipitated with antibodies to TRAF3 and analysed by immunoblotting with antibodies to Flag, TRIF and TRAF3. **e**, RAW264.7 cells were stimulated with CpG-DNA, lysates prepared and after immunoprecipitation with indicated antibodies analysed by either immunoblotting (upper panel) or kinase assays using GST-IRF7 or GST-IκBα as substrates (lower panels).

but both LPS and poly(I:C) (synthetic double-stranded RNA) induced association of TRAF3 with TRIF, suggesting that TRAF3 is also involved in the TRIF-dependent pathway (Fig. 2d). The kinetics of CpG-DNA-induced association of TRAF3 with MyD88 paralleled those of TBK1, but not IKK, activation (Fig. 2e). Together, the data show that TRAF3 is part of MyD88- and TRIF-dependent signalling complexes.

Heretofore, TRAF3 was not implicated in TLR signalling and its biological function has been enigmatic. *Traf3*^{-/-} mice die shortly after birth, exhibiting progressive hypoglycaemia and runting¹². Wild-type mice reconstituted with TRAF3-deficient bone marrow are viable and we used them as a source for *Traf3*^{-/-} myeloid cells. *Traf3*^{-/-} bone marrow-derived macrophages (BMDMs) and wild-type counterparts were stimulated with CpG-DNA and LPS. Only small differences in JNK1/2 and p38 phosphorylation were observed, and the kinetics of IκBα degradation were comparable (Supplementary Fig. 3a). However, CpG-DNA- and LPS-induced IL-12 and IL-6 expression was strongly elevated in *Traf3*^{-/-} BMDMs, whereas IL-10 production was undetectable (Fig. 3a and Supplementary Fig. 3b). Similar alterations regarding all IL-12 family members and IL-10 were observed at the messenger RNA level, whereas other genes were not affected (Supplementary Figs 3c and 4). Upregulation of the pro-inflammatory cytokines IL-12 and IL-6 was largely due to defective production of IL-10, an anti-inflammatory cytokine¹³, as exogenous IL-10 counteracted this effect (Supplementary Fig. 3d). Wild-type and *Traf3*^{-/-} cells also differed in the regulation of IFN and IFN-dependent genes. Induction of bioactive IFN by CpG-DNA, LPS and R-848, an imidazoquinoline (TLR7)¹⁴, was defective in *Traf3*^{-/-} BMDMs and in *Traf3*^{-/-} Flt3 ligand (Flt3L)-induced dendritic cells, the wild-type counterparts of which are known to produce high levels of IFN (Fig. 3b, c and Supplementary Fig. 3e)¹⁵.

Consistent with the observation that LPS activates IFN via TRIF^{4,5}, poly(I:C), which signals solely through TRIF^{4,5}, failed to induce IFN and IL-10 in *Traf3*^{-/-} BMDMs, yet induced over-production of IL-12 (Fig. 3d and Supplementary Fig. 3f). Vesicular stomatitis virus (VSV), which can activate cells via TLR7 and MyD88 (ref. 16), failed to induce IFN in *Traf3*^{-/-} BMDMs (Fig. 3e), which were also more susceptible to viral-induced killing than wild-type counterparts (data not shown).

We compared the function of TRAF3 to TRAF6. In *Traf6*^{-/-} BMDMs, CpG-DNA and Pam₃Cys did not induce MAPK phosphorylation or IκBα degradation, but the response to poly(I:C) proceeded normally (Fig. 4a). Thus, TRAF6 is critical for MyD88

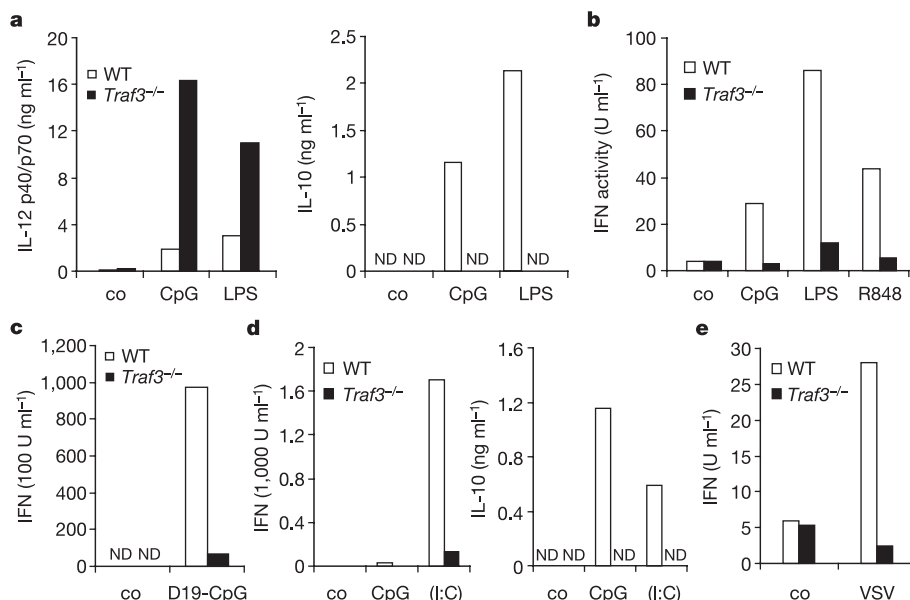


Figure 3 | TRAF3 is required for TLR-induced IFN and IL-10 production. **a**, BMDMs were stimulated for 16 h with CpG-DNA or LPS, and secretion of IL-12 p40/p70 and IL-10 was determined by ELISA. **b**, BMDMs were stimulated for 6 h with CpG-DNA, LPS or R-848 and supernatants were bioassayed for IFN activity. **c**, Flt3L-induced dendritic cells were stimulated for 24 h with D19-CpG-DNA and supernatants were bioassayed for IFN activity. **d**, BMDMs were stimulated with CpG-DNA or poly(I:C) for 16 h, and IFN and IL-10 secretion was determined as above. **e**, BMDMs were exposed or not to VSV (multiplicity of infection = 1) for 2 h. After 18 h, IFN activity in culture supernatants was determined. Experiments shown were repeated at least three times and representative results are presented. co, control; ND, not determined.

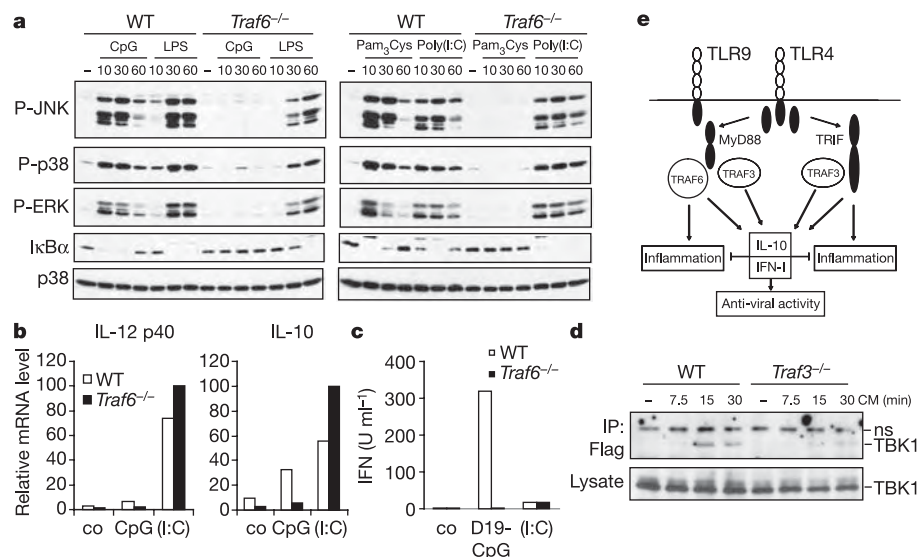


Figure 4 | Divergent TRAF6 and TRAF3 functions provide specificity in TLR signalling. **a**, BMDMs were stimulated with the indicated pathogen-associated molecular patterns and analysed for MAPK phosphorylation and IκBα degradation by immunoblotting. **b**, BMDMs were stimulated with CpG-DNA or poly(I:C). After 6 h the indicated mRNAs were quantified by Q-PCR. **c**, Ftl3L-induced dendritic cells were stimulated for 6 h with D19-CpG-DNA and IFN secretion was analysed. Experiments shown (**b**, **c**)

were repeated at least three times and representative results are presented. **d**, HoxB8-immortalized myeloid progenitors expressing Flag-MyD88-GyrB were stimulated with coumermycin, and lysates were prepared and analysed by immunoblotting with anti-TBK1 antibodies, either directly or after immunoprecipitation with antibodies to Flag. ns, non-specific. **e**, Model showing the different functions of TRAF3 and TRAF6 in TLR-mediated immune cell activation.

signalling, but not TRIF signalling. Consistent with this interpretation, the LPS response was only partially reduced in the absence of TRAF6 (Fig. 4a). Accordingly, TLR9-induced gene expression was dependent on TRAF6, whereas the response to poly(I:C) was TRAF6 independent, both in BMDMs and Ftl3L-induced dendritic cells (Fig. 4b, c and Supplementary Fig. 5a, b). Notably, many genes, including IL-10, IFN-β and IL-12 p40 were normally induced by LPS in *Traf6*^{-/-} macrophages (Supplementary Fig. 5c). Taken together, these results strongly suggest that LPS regulates type I IFN and IL-10 via the TRIF-dependent pathway and that this pathway relies on TRAF3 rather than TRAF6.

Two similar protein kinases, TBK1/NAK and IKKi/IKKε, are involved in type I IFN induction^{17,18}. Notably, the TBK1 interactor TANK/i-TRAF¹⁹ is also a TRAF2/3-interacting protein²⁰. To examine the role of TRAF3 in recruiting TBK1 to TIR complexes, we immortalized *Traf3*^{-/-} fetal liver myeloid progenitors with a retrovirus expressing the HoxB8 oncogene²¹. These cells were transduced with the MyD88-GyrB fusion construct. Coumermycin activated typical TLR signalling pathways (Supplementary Fig. 6a) and induced efficient recruitment of TBK1 to MyD88 in wild-type cells, but not in *Traf3*^{-/-} cells (Fig. 4d). Residual TBK1 recruitment in the absence of TRAF3 may be mediated by TRAF6. Immortalized *Traf3*^{-/-} cells also exhibited defective IFN and IL-10 production, which was corrected by re-introduction of TRAF3 (Supplementary Fig. 6b–d). Notably, IRAK1, which is also involved in IFN expression²², was equally degraded upon stimulation of wild-type and *Traf3*^{-/-} cells (data not shown), suggesting that it functions upstream of the TRAFs.

On the basis of these results we propose that TLR-induced MyD88 dimerization and IRAK4/IRAK1 activation initiates recruitment of TRAF6 and TRAF3 (Fig. 4e). The TRAF6-dependent pathway engages MAPKs and IKK, resulting in activation of transcription factors such as AP-1 and NF-κB that participate in the induction of pro-inflammatory cytokines. TRAF3, by contrast, is dispensable for MAPK and IKK activation or production of pro-inflammatory cytokines. TRAF3, however, is instrumental for recruitment of TBK1 and production of type I IFNs and IL-10. As a result

of IL-10 deficiency, *Traf3*^{-/-} myeloid cells overproduce the pro-inflammatory cytokines IL-12 and IL-6. TRAF3 has a much more important role than TRAF6 in TRIF-dependent signalling to the IFN and IL-10 genes, probably by marshalling TBK1 (and possibly IKKi) into the TIR signalling complex leading to activation of IRF3 and IRF7. The ratio of TRAF6 versus TRAF3 recruitment is likely to contribute to the biological specificity and consequences of TLR signalling.

METHODS

Mice and cell culture. *Traf6*^{-/-} mice were provided by J. Inoue⁶. Fetal liver transplantation has been described previously¹². Bone marrow was collected from mice 6–8 weeks after reconstitution and BMDMs or Ftl3L-induced dendritic cells were generated by using either L-cell conditioned medium or 50 ng ml⁻¹ Ftl3L (R&D Systems) as described^{23,24}. Stable RAW264.7 transfectants were generated and cultured as described²⁵. Fetal liver myeloid progenitors were immortalized using an oestrogen-responsive ER-HoxB8 oncogene in the presence of 1 μM oestrogen and GM-CSF²¹.

Reagents and plasmids. Antibodies are described in Supplementary Methods. Full-length mouse MyD88 was amplified by polymerase chain reaction with reverse transcription (RT-PCR) using cDNA from BMDMs, and fused to a triple Flag tag and subunit B of *E. coli* DNA gyrase⁷. Where indicated, a TAP tag was fused to the carboxy terminus of GyrB⁸. CpG-DNA refers to oligonucleotides (ODN) 1668 (1 μM; TIB Molbiol, TCCATGACGTTCTGTATGCT); D19-CpG refers to CpG-ODN D19 (1 μM; Invitrogen, GGTGC ATCGATGCAGGGGGG). Nucleotides with a phosphothioate backbone are in italics. Other agonists used are: LPS, 10 ng ml⁻¹ (*E. coli* 0127:B8, Sigma-Aldrich); poly(I:C), 30 μg ml⁻¹ (GE Healthcare); R-848, 300 nM (GLS); Pam₃Cys (tripalmitoyl cysteinyl lipopeptide), 1 μg ml⁻¹ (EMC Microcollections); coumermycin, 0.1 μM (Sigma-Aldrich); and IL-1β, 30 ng ml⁻¹ (PeproTech).

Analysis of gene expression. Total cellular RNA was prepared using TRIzol (Invitrogen) and analysed by real-time quantitative polymerase chain reaction (Q-PCR)²³. Primer sequences are available upon request. All values were normalized to the level of cyclophilin mRNA. To determine IFN bioactivity, a VSV-based bioassay was used⁴. All values represent means of duplicates. Reference murine IFN-α (NIAID/NIH) served as a standard. All ELISAs (IL-12 p40/p70, IL-6, IL-10) were from BD Biosciences.

Protein purification, analysis and mass spectrometry. TAP has been described previously⁸. Variations to the original protocol are described in Supplementary Methods. After TAP, samples were concentrated using

Biomax-5K centrifuge filters (Millipore), separated on a 4–12% Novex Bis-Tris gel (Invitrogen) and stained with colloidal Coomassie blue (Invitrogen). Protein bands were subjected to in-gel trypsin digestion and analysed by LC-MS/MS with a QSTAR-Pulsar quadrupole time-of-flight instrument (ABI-MDS-SCIEX)⁹. Immunoprecipitations and kinase assays were performed as described²⁵. Substrate proteins used in *in vitro* kinase assays (I κ B α ^{1–54}, IRF7^{409–457}) were expressed and purified as glutathione S transferase (GST) fusions.

Received 29 July; accepted 24 October 2005.

Published online 23 November 2005.

- Beutler, B. Inferences, questions and possibilities in Toll-like receptor signalling. *Nature* **430**, 257–263 (2004).
- Yamamoto, M., Takeda, K. & Akira, S. TIR domain-containing adaptors define the specificity of TLR signalling. *Mol. Immunol.* **40**, 861–868 (2004).
- Hemmi, H. *et al.* A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740–745 (2000).
- Hoebe, K. *et al.* Identification of Lps2 as a key transducer of MyD88-independent TIR signalling. *Nature* **424**, 743–748 (2003).
- Yamamoto, M. *et al.* Role of adaptor TRIF in the MyD88-independent toll-like receptor signalling pathway. *Science* **301**, 640–643 (2003).
- Gohda, J., Matsumura, T. & Inoue, J. Cutting edge: TNFR-associated factor (TRAF) 6 is essential for MyD88-dependent pathway but not toll/IL-1 receptor domain-containing adaptor-inducing IFN- β (TRIF)-dependent pathway in TLR signalling. *J. Immunol.* **173**, 2913–2917 (2004).
- Farrar, M. A., Alberol, I. & Perlmutter, R. M. Activation of the Raf-1 kinase cascade by coumermycin-induced dimerization. *Nature* **383**, 178–181 (1996).
- Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
- Blagoev, B., Ong, S. E., Kratchmarova, I. & Mann, M. Temporal analysis of phosphotyrosine-dependent signalling networks by quantitative proteomics. *Nature Biotechnol.* **22**, 1139–1145 (2004).
- Suzuki, N. *et al.* Severe impairment of interleukin-1 and Toll-like receptor signalling in mice lacking IRAK-4. *Nature* **416**, 750–756 (2002).
- Yamamoto, M. *et al.* Essential role for TIRAP in activation of the signalling cascade shared by TLR2 and TLR4. *Nature* **420**, 324–329 (2002).
- Xu, Y., Cheng, G. & Baltimore, D. Targeted disruption of TRAF3 leads to postnatal lethality and defective T-dependent immune responses. *Immunity* **5**, 407–415 (1996).
- Fiorino, D. F., Zlotnik, A., Mosmann, T. R., Howard, M. & O'Garra, A. IL-10 inhibits cytokine production by activated macrophages. *J. Immunol.* **147**, 3815–3822 (1991).
- Hemmi, H. *et al.* Small anti-viral compounds activate immune cells via the TLR7 MyD88-dependent signalling pathway. *Nature Immunol.* **3**, 196–200 (2002).
- Kawai, T. *et al.* Interferon- α induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nature Immunol.* **5**, 1061–1068 (2004).
- Lund, J. M. *et al.* Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc. Natl Acad. Sci. USA* **101**, 5598–5603 (2004).
- Fitzgerald, K. A. *et al.* IKK ϵ and TBK1 are essential components of the IRF3 signalling pathway. *Nature Immunol.* **4**, 491–496 (2003).
- Hemmi, H. *et al.* The roles of two I κ B kinase-related kinases in lipopolysaccharide and double stranded RNA signalling and viral infection. *J. Exp. Med.* **199**, 1641–1650 (2004).
- Pomerantz, J. L. & Baltimore, D. NF- κ B activation by a signalling complex containing TRAF2, TANK and TBK1, a novel IKK-related kinase. *EMBO J.* **18**, 6694–6704 (1999).
- Cheng, G. & Baltimore, D. TANK, a co-inducer with TRAF2 of TNF- and CD 40L-mediated NF- κ B activation. *Genes Dev.* **10**, 963–973 (1996).
- Knoepfler, P. S., Sykes, D. B., Pasillas, M. & Kamps, M. P. HoxB8 requires its Pbx-interaction motif to block differentiation of primary myeloid progenitors and of most cell line models of myeloid differentiation. *Oncogene* **20**, 5440–5448 (2001).
- Uematsu, S. *et al.* Interleukin-1 receptor-associated kinase-1 plays an essential role for Toll-like receptor (TLR)7- and TLR9-mediated interferon- α induction. *J. Exp. Med.* **201**, 915–923 (2005).
- Park, J. M., Greten, F. R., Li, Z. W. & Karin, M. Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science* **297**, 2048–2051 (2002).
- Hochrein, H. *et al.* Herpes simplex virus type-1 induces IFN- α production via Toll-like receptor 9-dependent and -independent pathways. *Proc. Natl Acad. Sci. USA* **101**, 11416–11421 (2004).
- Hacker, H. *et al.* Cell type-specific activation of mitogen-activated protein kinases by CpG-DNA controls interleukin-12 release from antigen-presenting cells. *EMBO J.* **18**, 6973–6982 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Inoue for *Traf6*-deficient mice, M. Farrar for GyrB constructs, G. Oganessian, S. K. Saha and G. Cheng for providing cells from TRAF3-deficient mice, T. Maniatis for anti-TBK1 antibodies and L. Pfeffer for L929 cells. H.H. and V.R. were supported by fellowships from the Deutsche Forschungsgemeinschaft (DFG). Work in the laboratories of M.P.K., E.R. and M.K. was supported by grants from the National Institutes of Health. M.K. is an American Cancer Society Research Professor.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.H. (Hans.Haecker@stjude.org).

LETTERS

Critical role of TRAF3 in the Toll-like receptor-dependent and -independent antiviral response

Gagik Oganessian^{1,2*}, Supriya K. Saha^{1,2*}, Beichu Guo¹, Jeannie Q. He¹, Arash Shahangian^{1,2}, Brian Zarnegar¹, Andrea Perry¹ & Genhong Cheng^{1,3}

Type I interferon (IFN) production is a critical component of the innate defence against viral infections¹. Viral products induce strong type I IFN responses through the activation of Toll-like receptors (TLRs) and intracellular cytoplasmic receptors such as protein kinase R (PKR)^{2–12}. Here we demonstrate that cells lacking TRAF3, a member of the TNF receptor-associated factor family, are defective in type I IFN responses activated by several different TLRs. Furthermore, we show that TRAF3 associates with the TLR adaptors TRIF and IRAK1, as well as downstream IRF3/7 kinases TBK1 and IKK- ϵ , suggesting that TRAF3 serves as a critical link between TLR adaptors and downstream regulatory kinases important for IRF activation. In addition to TLR stimulation, we also show that TRAF3-deficient fibroblasts are defective in their type I IFN response to direct infection with vesicular stomatitis virus, indicating that TRAF3 is also an important component of TLR-independent viral recognition pathways. Our data demonstrate that TRAF3 is a major regulator of type I IFN production and the innate antiviral response.

TNF receptor-associated factor (TRAF) family members, comprising TRAF1–6, are proposed to be adaptor molecules linking upstream receptor signals to downstream gene activation. Although multiple TRAFs are required for NF- κ B and JNK activation by the TNF receptor members, including CD40 and TNFR2, TRAF6 was thought to be the sole TRAF family adaptor involved in interleukin (IL)-1 receptor and TLR signalling^{13,14}. The function of TRAF3, a particularly elusive member of the TRAF family, has not yet been clearly defined¹⁵. TRAF3 was initially used as bait in a yeast two-hybrid screen several years ago to identify TRAF family member-associated NF- κ B activator (TANK), which is an adaptor that was later found to associate with the IRF3 and IRF7 kinases TBK1 and IKK- ϵ ^{16–19}. NAK-associated protein 1 (NAP1), a TANK homologue, has recently been suggested to be involved in IRF3 activation²⁰. Because the activation of IRF3/7 is an important event for type I IFN induction, this raises the possibility that TRAF3 could be directly involved in the IFN pathway^{11,21,22}. Because TRAF3 deficiency results in early postnatal lethality, we generated bone marrow-derived macrophages (BMMs) from irradiated C57BL/6 mice, reconstituted with *Traf3*^{+/+} and *Traf3*^{-/-} fetal liver cells¹⁵. To study the involvement of TRAF3 in TLR signalling, *Traf3*^{+/+} and *Traf3*^{-/-} BMMs were stimulated with poly(I:C), a synthetic double-stranded RNA, and lipopolysaccharide (LPS), the ligands for TLR3 and TLR4, respectively. Notably, the induction of *Ifnb* (which encodes IFN- β) and subsequent IFN target genes in response to poly(I:C) and LPS was markedly reduced in TRAF3-deficient BMMs (Fig. 1a and Supplementary Fig. 3a), whereas no defect was observed in the activation of NF- κ B or the induction of inflammatory-related

genes such as *Ikb α* and *KC* in *Traf3*-deficient BMMs (Fig. 1a and Supplementary Fig. 2d). *Traf3*^{-/-} BMMs were not inherently defective in their ability to produce type I IFNs, as infection with *Listeria monocytogenes* resulted in potent IFN- β induction in these cells (Supplementary Fig. 4). Furthermore, conditioned media from poly(I:C)-treated wild-type BMMs potently blocks the replication of both murine gammaherpesvirus 68 (MHV68) and vesicular stomatitis virus (VSV) in NIH3T3 cells (Fig. 1b and Supplementary Fig. 3)¹¹. In contrast, conditioned media from poly(I:C)-treated *Traf3*^{-/-} BMM cultures had no effect on viral replication, demonstrating that TRAF3 is critical for the ability of TLR-stimulated cells to inhibit viral replication.

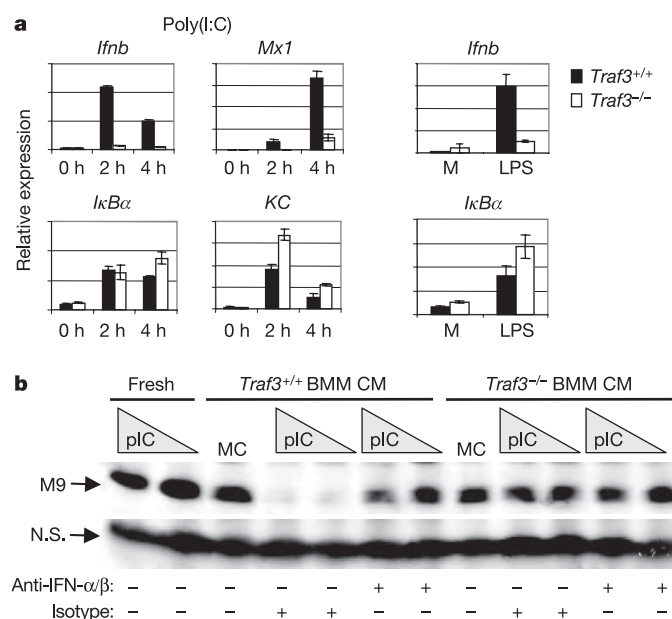


Figure 1 | Induction of the type I IFN response by TLR3 and TLR4 requires TRAF3. **a**, RNA from *Traf3*^{+/+} or *Traf3*^{-/-} BMMs stimulated with poly(I:C) (1 μ g ml⁻¹) or LPS (20 ng ml⁻¹) was assayed for the expression of the indicated mRNA transcripts by Q-PCR. Error bars indicate $\pm$ s.d. between duplicates. M, media control. **b**, NIH3T3 cells were pre-treated for 3 h with conditioned media from BMMs stimulated with poly(I:C) (0.1 or 1 μ g ml⁻¹) in the presence of anti-IFN- α/β or nonspecific rabbit IgG (20 μ g ml⁻¹). Cells were then infected with MHV68. Twenty-four hours after infection, MHV68 replication was assayed by immunoblotting for MHV68 M9 protein. MC, media control; N.S., non-specific band.

¹Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, 609 Charles E. Young Dr. East, Los Angeles, California 90095, USA.

²Medical Scientist Training Program, David Geffen School of Medicine at UCLA, 23-385 Center for the Health Sciences, Los Angeles, California 90095, USA. ³UCLA's Jonsson Comprehensive Cancer Center, 8-240 Factor Building, Los Angeles, California 90095, USA.

*These authors contributed equally to this work.

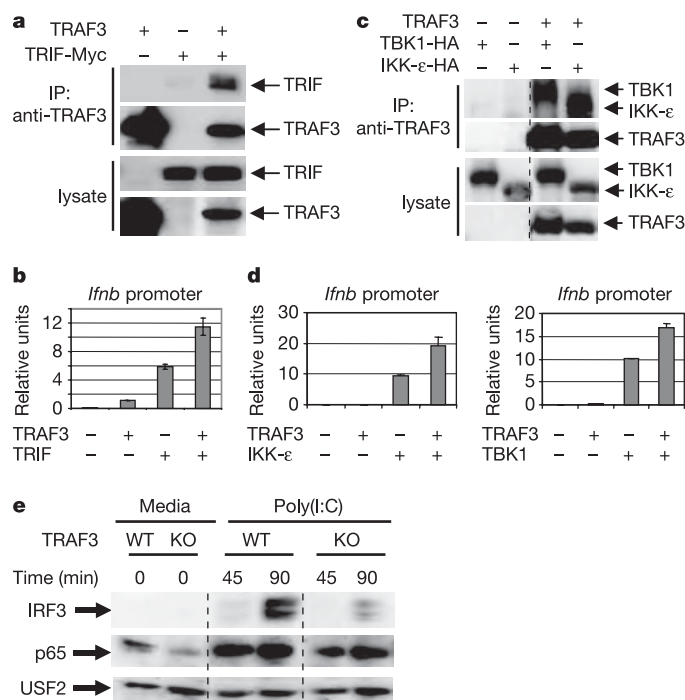


Figure 2 | TRAF3 associates with IFN pathway mediators and is required for TLR3-induced IRF3 activation. **a**, TRAF3 associates with Myc-TRIF when co-expressed in 293T cells by immunoprecipitation assay. **b**, TRAF3 synergizes with TRIF when co-expressed in 293T cells to activate *Ifnb* promoter by luciferase assay. **c**, TRAF3 associates with both haemagglutinin (HA)-TBK1 and HA-IKK-ε when co-expressed in 293T cells by immunoprecipitation assay. **d**, TRAF3 synergizes with both HA-TBK1 and HA-IKKε when co-expressed in 293T cells to activate *Ifnb* promoter by luciferase assay. **e**, Nuclear extracts from poly(I:C)-stimulated BMMs were monitored for the appearance of IRF3 and the p65 subunit of NF-κB by immunoblotting. KO, knockout; WT, wild type. Error bars in **b** and **d** indicate $\pm$ s.d. between duplicates.

TRIF is a critical adaptor molecule required for IFN production after poly(I:C) and LPS stimulation^{23,24}. Our finding that *Traf3*^{-/-} cells were defective in poly(I:C)- and LPS-mediated *Ifnb* upregulation suggested that TRAF3 might serve as an adaptor molecule linking TRIF to downstream IRF3 activation. In support of this hypothesis, we found that TRAF3 both associated with TRIF and significantly enhanced TRIF-mediated *Ifnb* promoter activity (Fig. 2a, b). Similarly, TRAF3 also associated with both TBK1 and IKK-ε, and synergized with both IRF3 kinases to enhance their activation of the *Ifnb* promoter (Fig. 2c, d). To determine whether TRAF3 is specifically involved in the activation and nuclear translocation of IRF3, nuclear extracts from poly(I:C)-treated wild-type and *Traf3*^{-/-} BMMs were analysed for the presence of IRF3. Strong IRF3 nuclear translocation was detected in poly(I:C)-stimulated wild-type BMMs compared to only trace amounts of nuclear IRF3 in TRAF3-deficient cells (Fig. 2e). Together, these observations demonstrate that TRAF3 is a critical adaptor molecule for TRIF-mediated IRF3 activation and IFN-β production.

Plasmacytoid dendritic cells (pDCs) are thought to be the major producers of type I IFNs *in vivo*²⁵. TLR9 and TLR7 activation induces potent MyD88- and IRAK1-dependent type I IFN production in pDCs^{4–10,26}. To address whether TRAF3 is also involved in TLR7/9-dependent type I IFN induction, Flt3-ligand-derived dendritic cells were generated from *Traf3*^{+/+} and *Traf3*^{-/-} bone marrow and stimulated with CpG and R848, synthetic ligands for TLR9 and TLR7, respectively. IFN-α production was severely defective in the TRAF3-deficient cells treated with either TLR9 or TLR7 agonist (Fig. 3a and Supplementary Fig. 5). These findings were further confirmed in purified pDCs (Fig. 3b, c and Supplementary Figs 6 and 7). Bone marrow cells for these experiments were isolated from *Traf3*^{+/+}*p100*^{-/-} and *Traf3*^{-/-}*p100*^{-/-} mice, which survive into adulthood (J. He *et al.*, manuscript in preparation).

It has recently been demonstrated that IRAK1 is specifically required for TLR9-mediated IFN-α production, whereas it is dispensable for TLR9-induced inflammatory responses²⁶. Our results reveal a similar phenotype in TRAF3-deficient cells. In fact, TRAF3 associated with IRAK1 when co-expressed in 293T cells and strongly synergized with IRAK1 and IRF7 to activate *Ifna4* promoter in a reporter assay (Fig. 3d, e). Thus, TRAF3 might function in a complex with IRAK1 to phosphorylate directly IRF7, or facilitate the

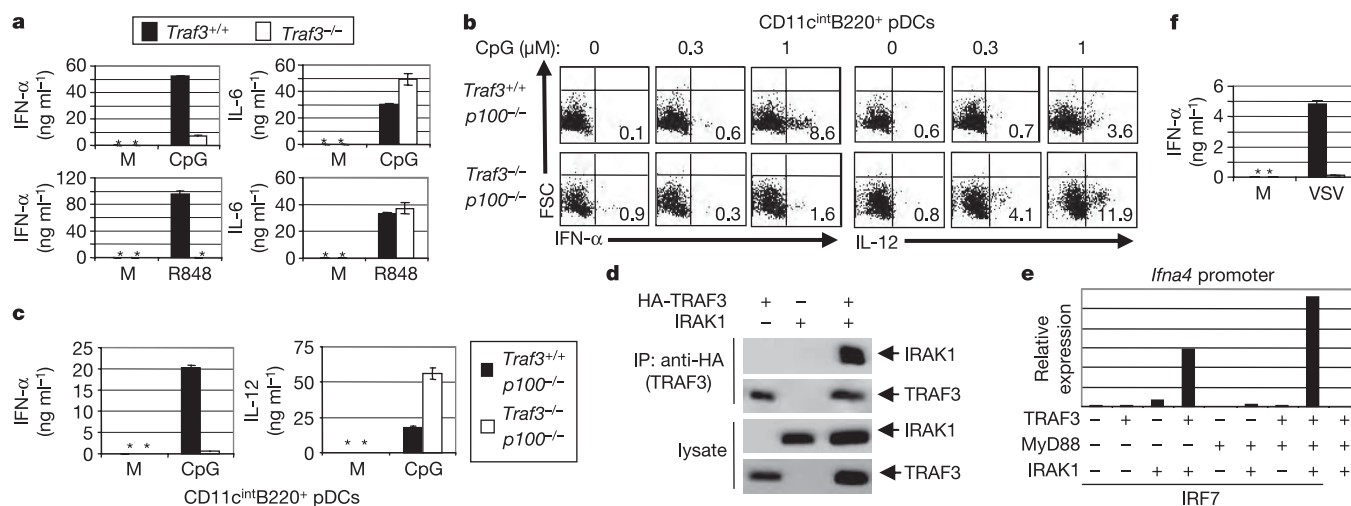


Figure 3 | TRAF3 is required for TLR7/9-dependent type I IFN production. **a**, **f**, Twenty-four-hour supernatants from Flt3-ligand-derived dendritic cells, stimulated with CpG D19 (1 μM), R848 (200 nM) (**a**) or infected with VSV (multiplicity of infection (MOI) of 1) (**f**), were analysed for the levels of IFN-α or IL-6 by ELISA (asterisk indicates not detected). M, media control. **b**, *p100*^{-/-}*Traf3*^{+/+} and *p100*^{-/-}*Traf3*^{-/-} Flt3-ligand-derived dendritic cells were stimulated with CpG D19 for 8 h, and

cells were analysed for the presence of intracellular IFN-α and IL-12. **c**, CD11c^{int}B220⁺ pDCs were purified by cell sorting from Flt3-ligand-derived dendritic cells and stimulated with CpG D19 for 24 h. IL-12 and IFN-α levels were analysed by ELISA. **d**, **e**, TRAF3 associates and synergizes with IRAK1 when co-expressed in 293T cells. Error bars in **a**, **c** and **f** indicate $\pm$ s.d. between duplicates.

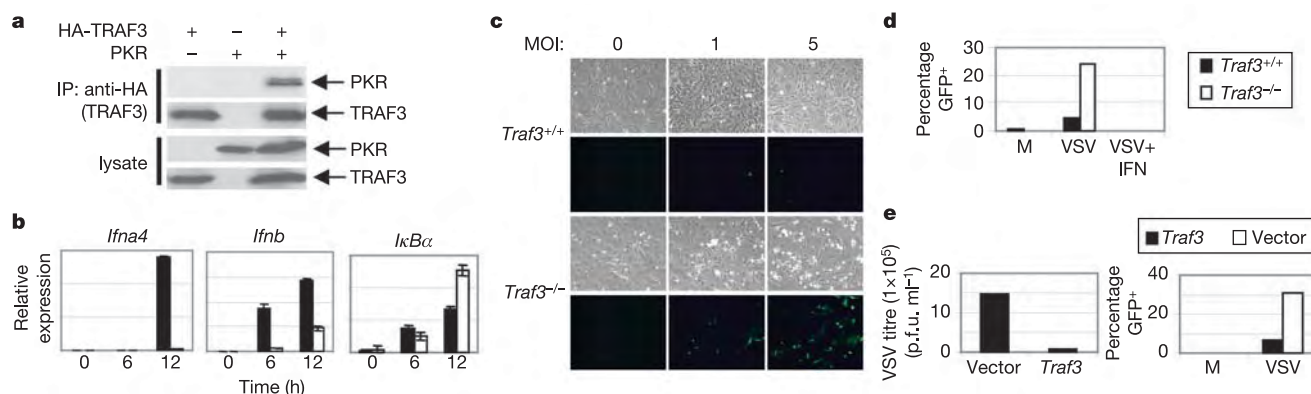


Figure 4 | Involvement of TRAF3 in TLR-independent antiviral responses. **a**, TRAF3 associates with PKR when co-expressed in 293T cells by immunoprecipitation assay. **b**, MEFs were infected with VSV and analysed for expression of the indicated genes by Q-PCR. *Traf3*^{+/+}, filled bars; *Traf3*^{-/-}, empty bars. Error bars indicate $\pm$ s.d. between duplicates. **c**, **d**, MEFs were infected with VSV expressing GFP (multiplicity of infection

of 1) in the presence or absence of IFN- β (100 U ml⁻¹). Twenty-four hours later, GFP expression was visualized by fluorescent microscopy and quantified by flow cytometry. **e**, *Traf3*^{-/-} MEFs were reconstituted with TRAF3, infected with VSV, and viral replication was quantified by plaque assays and flow cytometry. TRAF3 expression in reconstituted cells was confirmed by immunoblotting (data not shown).

activation of another downstream IRF7 kinase (Supplementary Fig. 1).

Defects in TLR-mediated type I IFN induction in TRAF3-deficient cells prompted us to investigate whether TRAF3 is also involved in viral-mediated type I IFN production. To address this, *Traf3*^{+/+} and *Traf3*^{-/-} Flt3-ligand-derived dendritic cells were infected with VSV. As shown in Fig. 3f, *Traf3*^{-/-} Flt3-ligand-derived dendritic cells were almost completely defective in the production of IFN- α in response to VSV. Induction of type I IFNs in pDCs by VSV is thought to proceed through a TLR7- and MyD88-dependent pathway¹⁰. In contrast to pDCs, murine embryonic fibroblasts (MEFs) are thought to detect viral infection through cytoplasmic receptors such as PKR and RIG-I (refs 2, 27). PKR, a putative intracellular viral receptor, has a critical role in the induction of type I IFNs and antiviral responses to VSV infection in both MEFs and animal models^{2,28}. Interestingly, PKR has been shown to activate NF- κ B through a direct interaction with TRAF2 and TRAF5 (ref. 29). We found that PKR also associates with TRAF3 when expressed in 293T cells (Fig. 4a), suggesting a possible role for TRAF3 in PKR-dependent IFN production. In fact, type I IFN induction was defective in TRAF3-deficient MEFs in response to VSV infection (Fig. 4b). Accordingly, we observed increased susceptibility to VSV infection in TRAF3-deficient MEFs compared to wild-type controls (Fig. 4c, d). This defect is due to the loss of TRAF3 expression, as ectopically expressed TRAF3 in *Traf3*^{-/-} cells rescued their resistance to VSV infection (Fig. 4e).

Our findings indicate that TRAF3 is required for type I IFN production in response to both TLR activation and direct viral infection (Supplementary Fig. S1). Furthermore, we provide evidence that TRAF3 can associate with putative intracellular viral receptors, such as PKR, as well as TLR adaptors and downstream IRF3/7 kinases. These findings suggest that multiple TLR-dependent and TLR-independent pathways may converge on a common TRAF3 complex involved in the regulation of type I IFN production. Thus, we show that TRAF3 is a novel critical mediator of innate antiviral responses.

METHODS

Mice. C57BL/6 (Jackson Laboratories) mice aged 6–12 weeks were used as recipients in fetal liver transplantation experiments. Targeted disruption of the *Traf3* gene and the fetal liver reconstitutions have been described previously¹⁵. To obtain *Traf3*^{+/+} and *Traf3*^{-/-} *p100*^{-/-} cells, *Traf3*^{+/+} mice were crossed with *NF- κ B* *p100*^{-/-} mice. *Traf3*^{-/-} *p100*^{-/-} progeny live into adulthood (J. He et al., manuscript in preparation). All mice were maintained and bred under SPF conditions in the UCLA Life Sciences mouse facility. Bone marrow was harvested from reconstituted mice 6–8 weeks after reconstitution or from 8-day-old

Traf3^{+/+} and *Traf3*^{-/-} mouse pups and used for the generation of BMs or Flt3-ligand-derived dendritic cells.

Reagents. Flt3-ligand-derived dendritic cells and purified pDCs were stimulated with CpG D19 (ggTGCATCGATGCaggggG, where upper- and lower-case letters indicate bases with phosphodiester and phosphorothioate-modified backbones, respectively)⁶. BMs were stimulated with the phosphorothioate-modified CpG sequence TCCATGACGTTCCCTGACGTT (0.1 μ M). TBK1 and IKK- ϵ constructs were described previously³⁰. Full-length *Trif* was cloned from a murine complementary DNA library into a Myc-pCMV eukaryotic expression vector. Full-length *Irak1* and *Pkr* were cloned from a human cDNA library into a pcDNA3 vector.

Viral inhibition assays. Viral inhibition experiments were performed as described previously¹¹.

Quantitative PCR (Q-PCR). Q-PCR analyses were done using an iCycler thermocycler (Bio-Rad) as previously described¹¹. Primer sequences for *Ifna5*, *Ifnb*, *Mx1*, *IkBα*, *Isg15*, *Ifi204*, *L32*, *RANTES* (also known as *Ccl5*) and *IP10* are the same as those previously published^{11,30}. Primers for *KC*, *Irf3*, *Il12* and *Ifna4* are as follows: *KC*, forward 5'-CACTGCACCCAAACCGAAGT-3', reverse 5'-GGACAATTTCTGAACCAAGGG-3'; *Irf3*, forward 5'-ACATCTCCAA CAGCCAGCCTAT-3', reverse 5'-AGTCCATGTCCTCCACCAAGTC-3'; *Il12*, forward 5'-AAGTATTCAGTGTCTGCCAGGA-3', reverse 5'-TGCTTCC AACGCCAGTTCA-3'; *Ifna4*, forward 5'-CCTGTGTGATGCAGGAACC-3', reverse 5'-TCACCTCCCAGGCACAGA-3'. *L32* expression measurements were conducted in tandem with the gene of interest. All data are presented as relative expression units after normalization to the average *L32* value.

Received 29 July; accepted 13 October 2005.

Published online 23 November 2005.

- Isaacs, A. & Lindenmann, J. Virus interference. I. The interferon. By A. Isaacs and J. Lindenmann, 1957. *J. Interferon Res.* **7**, 429–438 (1987).
- Balachandran, S. et al. Essential role for the dsRNA-dependent protein kinase PKR in innate immunity to viral infection. *Immunity* **13**, 129–141 (2000).
- Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF- κ B by Toll-like receptor 3. *Nature* **413**, 732–738 (2001).
- Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S. & Reis e Sousa, C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* **303**, 1529–1531 (2004).
- Heil, F. et al. Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* **303**, 1526–1529 (2004).
- Hemmi, H., Kaisho, T., Takeda, K. & Akira, S. The roles of Toll-like receptor 9, MyD88, and DNA-dependent protein kinase catalytic subunit in the effects of two distinct CpG DNAs on dendritic cell subsets. *J. Immunol.* **170**, 3059–3064 (2003).
- Hemmi, H. et al. Small anti-viral compounds activate immune cells via the TLR7/MyD88-dependent signaling pathway. *Nature Immunol.* **3**, 196–200 (2002).
- Hochrein, H. et al. Herpes simplex virus type-1 induces IFN- α production via Toll-like receptor 9-dependent and -independent pathways. *Proc. Natl Acad. Sci. USA* **101**, 11416–11421 (2004).
- Krug, A. et al. Herpes simplex virus type 1 activates murine natural interferon-producing cells through toll-like receptor 9. *Blood* **103**, 1433–1437 (2004).

10. Lund, J. M. *et al.* Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc. Natl Acad. Sci. USA* **101**, 5598–5603 (2004).
11. Doyle, S. *et al.* IRF3 mediates a TLR3/TLR4-specific antiviral gene program. *Immunity* **17**, 251–263 (2002).
12. Tabeta, K. *et al.* Toll-like receptors 9 and 3 as essential components of innate immune defense against mouse cytomegalovirus infection. *Proc. Natl Acad. Sci. USA* **101**, 3516–3521 (2004).
13. Dempsey, P. W., Doyle, S. E., He, J. Q. & Cheng, G. The signaling adaptors and pathways activated by TNF superfamily. *Cytokine Growth Factor Rev.* **14**, 193–209 (2003).
14. Cao, Z., Xiong, J., Takeuchi, M., Kurama, T. & Goeddel, D. V. TRAF6 is a signal transducer for interleukin-1. *Nature* **383**, 443–446 (1996).
15. Xu, Y., Cheng, G. & Baltimore, D. Targeted disruption of TRAF3 leads to postnatal lethality and defective T-dependent immune responses. *Immunity* **5**, 407–415 (1996).
16. Fitzgerald, K. A. *et al.* IKK ϵ and TBK1 are essential components of the IRF3 signaling pathway. *Nature Immunol.* **4**, 491–496 (2003).
17. Cheng, G. & Baltimore, D. TANK, a co-inducer with TRAF2 of TNF- and CD 40L-mediated NF- κ B activation. *Genes Dev.* **10**, 963–973 (1996).
18. Sharma, S. *et al.* Triggering the interferon antiviral response through an IKK-related pathway. *Science* **300**, 1148–1151 (2003).
19. Pomerantz, J. L. & Baltimore, D. NF- κ B activation by a signaling complex containing TRAF2, TANK and TBK1, a novel IKK-related kinase. *EMBO J.* **18**, 6694–6704 (1999).
20. Sasai, M. *et al.* Cutting edge: NF- κ B-activating kinase-associated protein 1 participates in TLR3/Toll-IL-1 homology domain-containing adapter molecule-1-mediated IFN regulatory factor 3 activation. *J. Immunol.* **174**, 27–30 (2005).
21. Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **434**, 772–777 (2005).
22. Kawai, T. *et al.* Interferon- α induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nature Immunol.* **5**, 1061–1068 (2004).
23. Hoebe, K. *et al.* Identification of Lps2 as a key transducer of MyD88-independent TIR signalling. *Nature* **424**, 743–748 (2003).
24. Yamamoto, M. *et al.* Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science* **301**, 640–643 (2003).
25. Siegal, F. P. *et al.* The nature of the principal type 1 interferon-producing cells in human blood. *Science* **284**, 1835–1837 (1999).
26. Uematsu, S. *et al.* Interleukin-1 receptor-associated kinase-1 plays an essential role for Toll-like receptor (TLR)7- and TLR9-mediated interferon- α induction. *J. Exp. Med.* **201**, 915–923 (2005).
27. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
28. Chu, W. M. *et al.* JNK2 and IKK β are required for activating the innate response to viral infection. *Immunity* **11**, 721–731 (1999).
29. Gil, J. *et al.* TRAF family proteins link PKR with NF- κ B activation. *Mol. Cell. Biol.* **24**, 4502–4512 (2004).
30. Perry, A. K., Chow, E. K., Goodnough, J. B., Yeh, W. C. & Cheng, G. Differential requirement for TANK-binding kinase-1 in type I interferon responses to toll-like receptor activation and viral infection. *J. Exp. Med.* **199**, 1651–1658 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements G.O., S.K.S. and A.S. are supported by a UCLA Medical Scientist Training Program training grant. B.G. is supported by the Tumour Immunology Fellowship. B.Z. is supported by the Warsaw Fellowship. A.P. is supported by a H.H.M.I. pre-doctoral fellowship. G.C. is a Lymphoma and Leukemia Society Scholar, and part of this work was also supported by National Institutes of Health research grants.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.C. (genhongc@microbio.ucla.edu).

LETTERS

Generation of nuclear transfer-derived pluripotent ES cells from cloned *Cdx2*-deficient blastocysts

Alexander Meissner¹ & Rudolf Jaenisch¹

The derivation of embryonic stem (ES) cells by nuclear transfer holds great promise for research and therapy but involves the destruction of cloned human blastocysts. Proof of principle experiments have shown that 'customized' ES cells derived by nuclear transfer (NT-ESCs) can be used to correct immunodeficiency in mice¹. Importantly, the feasibility of the approach has been demonstrated recently in humans², bringing the clinical application of NT-ESCs within reach. Altered nuclear transfer (ANT) has been proposed as a variation of nuclear transfer because it would create abnormal nuclear transfer blastocysts that are inherently unable to implant into the uterus but would be capable of generating customized ES cells³. To assess the experimental validity of this concept we have used nuclear transfer to derive mouse blastocysts from donor fibroblasts that carried a short hairpin RNA construct targeting *Cdx2*. Cloned blastocysts were morphologically abnormal, lacked functional trophoblast and failed to implant into the uterus. However, they efficiently generated pluripotent embryonic stem cells when explanted into culture.

Survival of the normal embryo beyond implantation depends on the formation of the trophoblast lineage, the extra-embryonic lineage that forms the fetal-maternal interface within the placenta. The second embryonic lineage that forms, the inner cell mass (ICM), gives rise to all subsequent lineages in the embryo proper, and it is the ICM that, upon explanting in culture, gives rise to ES cells. The 'altered nuclear transfer' (ANT) concept³ is based on the premise that the inactivation of a gene crucial for trophoblast development will eliminate the potential to form the fetal-maternal interface, but will spare the ICM lineage. By genetically altering a somatic donor cell before nuclear transfer, one could generate cloned blastocysts that have no potential to develop beyond the blastocyst stage because no placenta could be formed. However, such cloned blastocysts could generate NT-ESCs derived from the ICM.

In this study we have performed a proof-of-principle experiment in mice to test the validity of the ANT approach and chose *Cdx2* as a candidate gene, as this gene encodes the earliest-known trophoblast-specific transcription factor that is activated in the 8-cell embryo and is essential for establishment and function of the trophoblast lineage^{4,5}. *Cdx2*-deficient blastocysts fail to maintain a blastocoel, lack epithelial integrity, dysregulate the ICM-specific transcription factors *Oct-4* and *Nanog*, and show increased cell death⁵. Importantly, *Cdx2*-deficient blastocysts are able to form an ICM and generate ES cells when explanted in tissue culture^{4,5}.

We selected for functional short hairpin (sh)RNAs against *Cdx2* as described in Supplementary Figs S1 and S2 (Supplementary Information). The experimental scheme, outlined in Fig. 1a, involved the introduction of a conditional *Cdx2* shRNA lentiviral vector (Fig. 1b) into primary tail-tip fibroblasts from neonatal F₁ mice (C57BL/6x129/SvJae). Green fluorescent protein (GFP)-positive *Cdx2*^{2Lox} tail-tip fibroblasts were selected and used as donors for

nuclear transfer. *Cdx2*-deficient blastocysts derived from the manipulated donor cells were tested for their potential to implant into the uterus and to generate pluripotent ES cells.

Of a total of 526 reconstructed oocytes, 350 formed pronuclei, of which 61 cleaved and developed into nuclear transfer morula/blastocysts. *Cdx2* knockdown nuclear transfer embryos showed no delay in developing to the early blastocyst stage compared to nuclear transfer embryos expressing a shRNA targeting *CD86* (data not shown). Figure 2a shows that GFP-positive *Cdx2*^{2Lox} nuclear transfer blastocysts did not express *Cdx2* as assessed by immunohistochemistry, in contrast to wild-type blastocysts (columns 1 and 2, Fig. 2a). Figure 2b shows that, when compared to control nuclear transfer blastocysts, *Cdx2*^{2Lox} nuclear transfer blastocysts were morphologically abnormal and failed to maintain a blastocoel cavity during *in vitro* cultivation, similar to previous results with *Cdx2* knockout blastocysts⁵. Using semi-quantitative polymerase chain reaction with reverse transcription (RT-PCR), we confirmed the

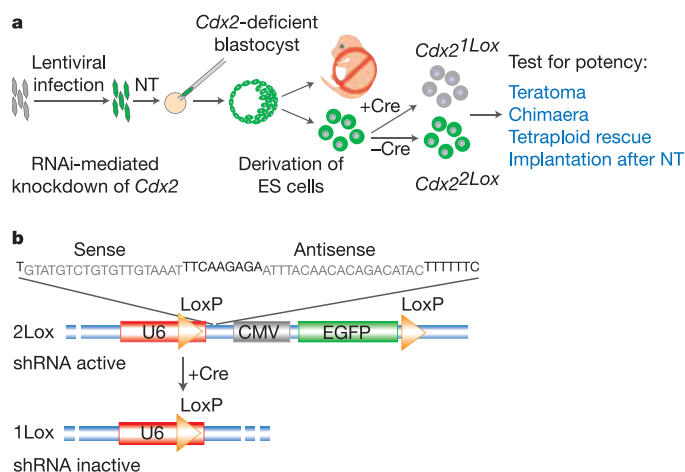


Figure 1 | Derivation of NT-ESCs from *Cdx2*-deficient blastocysts.

a, Primary tail-tip fibroblasts were infected with a conditional lentiviral RNA interference (RNAi) construct targeting *Cdx2* before nuclear transfer (NT). Blastocysts deficient for *Cdx2* were morphologically abnormal and unable to implant but gave rise to NT-ESCs. After initial expansion of the *Cdx2* knockdown NT-ESCs (*Cdx2*^{2Lox}) we used transient Cre expression to generate subclones (*Cdx2*^{1Lox}) with a deleted hairpin. To test the potency of ES lines before and after 'loop-out' we used teratoma formation, diploid and tetraploid blastocyst injections as well as nuclear transfer. **b**, The conditional RNAi system (pSicoR) has been described previously⁶. The shRNA, which targets nucleotides 1890–1908 located in the 3' UTR of *Cdx2*, was cloned into the conditional RNAi vector generating pSicoR-*Cdx2*^{2Lox}. This vector carries the *Cdx2* shRNA construct and an enhanced green fluorescence protein (EGFP) gene flanked by two LoxP sites (2Lox), which allows for Cre-mediated deletion of the shRNA and the EGFP sequences.

¹Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA.

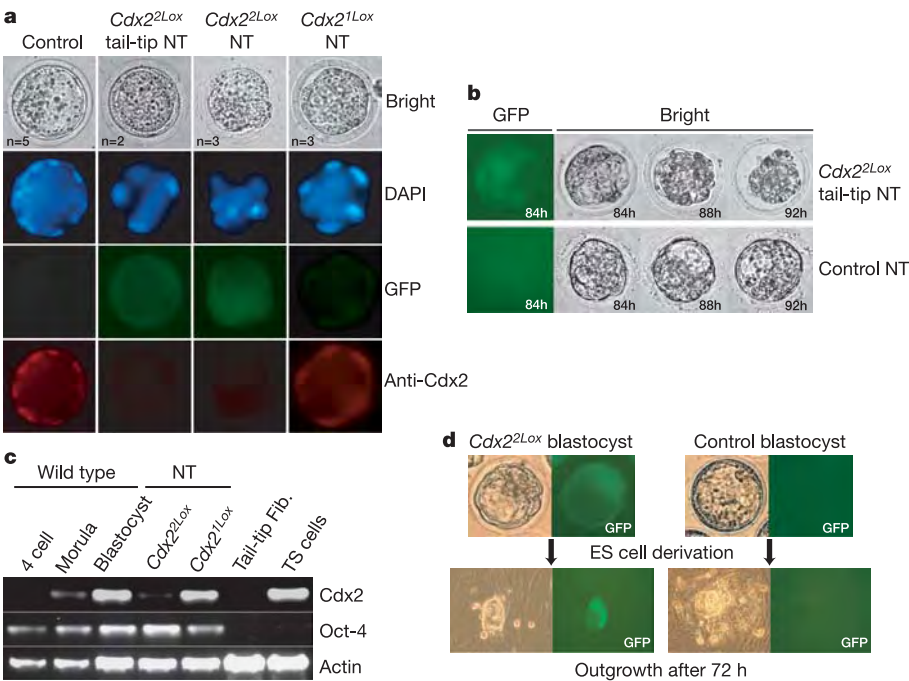


Figure 2 | *Cdx2*-deficient blastocysts and ES cell derivation. **a**, *Cdx2* immunostaining of day 3.5–4.5 wild-type and nuclear transfer blastocysts. The following donor cells were used for the nuclear transfer (from second column, left to right): *Cdx2*^{2Lox} tail-tip, *Cdx2*^{2Lox} ES cells, and *Cdx2*^{1Lox} ES cells. **b**, A typical *Cdx2*^{2Lox} tail-tip nuclear transfer blastocyst is shown 84 h after activation of the reconstructed oocytes. *Cdx2*-deficient blastocysts initially cavitated but failed to maintain the blastocoel and collapsed. Below, an expanded nuclear transfer blastocyst derived from control cells is shown. **c**, RT–PCR analysis of normal and *Cdx2*-deficient nuclear transfer pre-implantation morula/blastocysts. Four 4-cell embryos were pooled and RNA

was extracted for reverse transcription. All other samples were prepared from single morulae or blastocysts. Tail-tip fibroblasts (lane 6) express neither *Cdx2* nor *Oct-4*. Trophectoderm stem (TS) cells (lane 7) express *Cdx2*, but no *Oct-4*. A faint *Cdx2*-specific band, such as that seen in the blastocyst containing the shRNA construct targeting *Cdx2* shown in the figure, was detected in less than half of the tested embryos; most gave no signal in this test. **d**, Derivation of ES cells from *Cdx2*-deficient blastocysts. A *Cdx2*^{2Lox} tail-tip nuclear transfer-derived blastocyst with its initial outgrowth after 72 h (left) and a wild-type blastocyst (right) with its initial outgrowth are shown.

deficiency of *Cdx2* expression in *Cdx2*^{2Lox} nuclear transfer blastocysts, whereas control morulae and blastocysts showed robust *Cdx2* expression (Fig. 2c).

To assess whether *Cdx2* deficiency interfered with post-implantation development, *Cdx2*^{2Lox} nuclear transfer morulae/blastocysts were transferred into the uteri of pseudo-pregnant females. The uteri were removed at embryonic day (E)6.5 and examined for sites of implantation. Figure 3a shows no implantations in the uterus from a foster mother transplanted with five *Cdx2*^{2Lox} nuclear transfer blastocysts, in contrast to a uterus transplanted with five nuclear transfer control blastocysts that resulted in successful implantations (Fig. 3b). As summarized in Table 1, none of the *Cdx2*^{2Lox} nuclear transfer blastocysts formed visible implantation sites (0 out of 40), in contrast to control nuclear transfer blastocysts that were derived from fibroblasts carrying the *CD8* control shRNA (6 out of 15). In addition, no evidence for delayed implantation was obtained, as we failed to detect implantation sites at E7 or E8 in females transplanted with a total of 18 *Cdx2* knockdown nuclear transfer embryos (data not shown). These results demonstrate that nuclear transfer from donor fibroblasts carrying the pSicoR-*Cdx2*^{2Lox} virus resulted in morphologically abnormal, *Cdx2*-deficient nuclear transfer blastocysts that failed to implant upon transfer into foster mothers.

To investigate whether *Cdx2*-deficient blastocysts can generate ES cells upon explantation in culture, nuclear transfer *Cdx2*^{2Lox} blastocysts were transferred onto feeder cells. Whereas control nuclear transfer blastocysts formed trophoblastic outgrowths characteristic of the trophectoderm lineage, the *Cdx2*^{2Lox} nuclear transfer blastocysts failed to generate any trophoblast cells (Fig. 2d). Consistent with previous observations^{4,5}, *Cdx2*-deficient blastocysts generated ICM outgrowths that grew into stable, GFP-positive

nuclear transfer *Cdx2*^{2Lox} ES cell lines with an efficiency that was comparable to that of nuclear transfer blastocysts derived from wild-type fibroblasts (14% of explanted blastocysts). As a criterion for pluripotency, we tested the ability of the nuclear transfer *Cdx2*^{2Lox} ES cell lines to form chimaeras when injected into diploid blastocysts. The GFP-labelled cells contributed extensively to neonatal chimaeras (Fig. 3c, d) and formed high-grade postnatal chimaeras (Fig. 3i, summarized in Table 2) with high contributions to most tissues (Fig. 3e, f), with the notable exception of the intestine (Fig. 3g), which was entirely composed of *Cdx2*-positive cells derived from the host blastocyst (Fig. 3h). This is in agreement with previous reports, as it has been shown that *Cdx2* is required for normal development of the gastro-intestinal tract⁷. We further explored the developmental potency of the NT-ESCs using tetraploid complementation, which represents the most stringent test for pluripotency, as the resulting ‘ES mice’ are entirely composed of cells derived from the injected ES cells⁸. Consistent with previous results⁴, transfer of the *Cdx2*^{2Lox} ES

Table 1 | Survival of clones to blastocyst and post-implantation stage after nuclear transfer from different donor cells

Donor cells	pSicoR genotype	Number of clones with pseudo-pronuclei	Number of morulae/blastocysts	Number of implants at E6.5 (number of foster mothers)
Fibroblasts	<i>Cdx2</i> ^{2Lox}	211	40	0 (7)
Fibroblasts	<i>CD8</i> ^{2Lox}	76	15	6 (3)
ES cells	<i>Cdx2</i> ^{2Lox}	177	18	0 (4)
ES cells	<i>Cdx2</i> ^{1Lox}	199	22	11 (5)
ES cells	<i>CD8</i> ^{2Lox}	103	15	7 (3)

Shown are the number of reconstructed oocytes with pseudo-pronuclei after 5–6 h of activation. Morula/blastocyst transfers were done on day 3.5. pSicoR-*CD8*^{2Lox} fibroblasts carry a shRNA against *CD8* (ref. 6).

Table 2 | Developmental potential of ES cells deficient (*Cdx2*^{2Lox}) or proficient (*Cdx2*^{1Lox}) for *Cdx2* expression

Donor ES cells	Injected into 4N blastocysts			Injected into 2N blastocysts	
	Number injected	Embryos live at		Number injected	Number of chimaeras
		E14	Term		
<i>Cdx2</i> ^{2Lox}	112	0	0	65	12/23
<i>Cdx2</i> ^{1Lox}	36	ND	6	ND	ND

Thirty-three sites of implantation were detected in the tetraploid *Cdx2*^{2Lox} recipients, but no or only reabsorbed embryos (at 14.5) were present. Twelve of the 23 born diploid pups were chimaeras. ND, not done.

cells resulted in no live embryos at E14 (Table 2). These data indicate that nuclear transfer using pSicoR-*Cdx2*^{2Lox} fibroblasts generates abnormal blastocysts that are inherently unable to implant and grow into a fetus but are able to generate pluripotent ES cells that have a diminished developmental potency as compared to wild-type ES cells.

To assess whether NT-ESCs derived from *Cdx2*-deficient blastocysts could have the same pluripotency as wild-type ES cells, we investigated whether the block to normal developmental potential could be relieved by reversing the effects of the *Cdx2* gene knock-down. Normal *Cdx2* gene function was restored in *Cdx2*^{2Lox} ES cells by transient transfection of a Cre plasmid, resulting in the deletion of the *Cdx2* shRNA and *EGFP* marker gene (*Cdx2*^{2Lox} to *Cdx2*^{1Lox}; compare Fig. 1b), and rendering the cells *Cdx2* competent and GFP negative. Nuclear transfer from *Cdx2*^{1Lox} donor cells generated GFP-negative normal-appearing nuclear transfer blastocysts that expressed wild-type levels of *Cdx2*, as shown by immunostaining (Fig. 2a, right column) and RT-PCR (Fig. 2c, lane 5). To test whether deletion of the shRNA would restore pluripotency, the *Cdx2*^{1Lox} ES cells were injected into tetraploid blastocysts. As shown in Table 2, *Cdx2*^{1Lox} ES cells efficiently generated ES mice in contrast to the *Cdx2*^{2Lox} ES cells that were unable to give rise to ES mice. These results show that deletion of the *Cdx2* shRNA sequences creates ES cells that can generate all somatic tissues including normal intestinal cells, which cannot be derived from the *Cdx2*^{2Lox} parental ES cells (compare Fig. 3g, h). Finally, to test whether totipotency of *Cdx2*^{1Lox} ES cell nuclei was recovered, we transplanted *Cdx2*^{1Lox} blastocysts derived by nuclear transfer using *Cdx2*^{1Lox} donor ES cells into pseudo-pregnant foster mothers. As summarized in Table 1, normal-sized implants were detected at E6.5. These results confirm that *Cdx2* deficiency was responsible for the failure of clones to generate functional blastocysts and exclude other genetic alterations acquired during *in vitro* manipulation of the cells in the characteristic block to implantation. Most importantly, our data demonstrate that ES cells competent to generate all lineages can be derived from abnormal nuclear transfer blastocysts.

The ethical controversy surrounding nuclear transplantation arises from the necessity to destroy the reconstructed human blastocyst in order to obtain embryonic stem cells that can be used for biomedical research and therapy (for different view points see refs 9, 10). All available evidence is consistent with the conclusion that after nuclear transfer, the reconstructed embryos lack the potential to develop into normal human beings with any acceptable or practical efficiency¹¹. Despite the incompatibility with normal human development, the utility and promise of nuclear transfer is that embryonic stem cells derived by nuclear transfer have the same biological and molecular characteristics and the same therapeutic potential as those derived from fertilized embryos^{11,12}. Altered nuclear transfer further cripples an already compromised blastocyst and eliminates the developmental potential to implant into the uterus to establish the fetal-maternal connection³. The genetic manipulations of the somatic donor cells that are required to generate such an inherently abnormal blastocyst are simple and

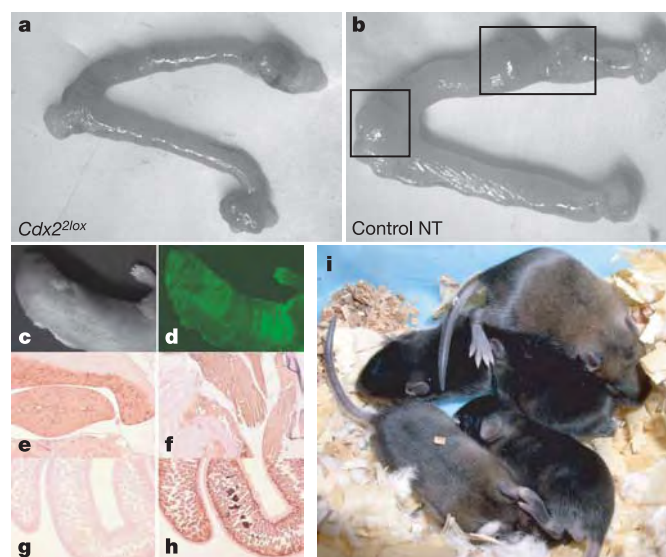


Figure 3 | *Cdx2*-deficient cells maintain developmental potential but are unable to implant after nuclear transfer. **a, b,** In each example shown, five nuclear transfer blastocysts were transferred at day 3.5 into the uterus of a day 2.5 pseudo-pregnant female. **a,** *Cdx2*-deficient blastocysts fail to implant. A representative uterus isolated at day 6.5 is shown. No deciduae were detectable from transplanted *Cdx2*-deficient blastocysts. **b,** Control nuclear transfer blastocysts showed normal implantation sites at day 6.5. **c,** Bright-field image of a postnatal *Cdx2*^{2Lox} ES chimaera. **d,** GFP signal indicates a contribution from *Cdx2*^{2Lox} ES cells. **e–g,** Histological sections and anti-GFP staining from a newborn *Cdx2*^{2Lox} chimaera. There was a contribution to the liver (endoderm; **e**) and muscle (mesoderm; **f**) but not to the intestine (**g**). **h,** Anti-*Cdx2* staining of the intestine shown in **g**. **i,** Coat colour contribution of *Cdx2*^{2Lox} ES cells. Recipient blastocysts have a C57BL/6 x DBA/2 F₁ background and the *Cdx2*^{2Lox} ES cells have a C57BL/6 x 129SvJae background. The presence of agouti (129/SvJae) fur indicates donor cell contribution. A litter with one wild type (black mouse below the top agouti), two low-contribution (middle) and two high-contribution chimaeras are shown.

straightforward. Our data indicate that the *Cdx2*-deficient blastocyst derived by nuclear transfer from a genetically engineered somatic cell is morphologically abnormal and lacks functional cells of the trophoblast lineage, consistent with previous results with embryos from mutant animals⁵. Because the gene is expressed before the blastocyst stage⁵, *Cdx2*-deficient clones are molecularly abnormal already at pre-blastocyst stages before an overtly abnormal phenotype becomes apparent. By reversing the *Cdx2* deficiency we demonstrate that fully competent ES cells can be derived from the inherently abnormal product of nuclear transfer using *Cdx2*-deficient donor cells.

If ANT was ever contemplated as an approach for the generation of human ES cells by nuclear transfer, the following issues need to be considered. (1) Although *CDX2* is expressed in the trophoblast of human blastocysts¹³ and derivatives of hES cells¹⁴ its expression pattern in the human fetus has not been determined and it is unknown whether it has an identical effect on placentation as in mouse. Because the effect of gene inhibition on human placentation cannot be directly tested, surrogate assays such as *in vitro* differentiation of ES cells are required to assess the effect of *CDX2* deficiency on human trophoblast development. (2) The use of retroviral vectors for gene transduction¹⁵ raises the possibility of insertional mutagenesis and the activation of oncogenes leading to leukaemia¹⁶. However, this probably does not represent a serious problem in ANT because, in contrast to the gene therapy trials using retroviral infection of bone marrow cells, viral integration into fibroblasts does not lead to a proliferative advantage and selective outgrowth of infected cells due to an activated oncogene. Also, because all nuclear

transfer ES cells are clonal, it would be easy to ensure by DNA analysis that proviral integration was not in the vicinity of an oncogene.

The results reported in this paper provide proof of principle that inhibition of genes important for trophoblast function can prevent placentalization without interfering with ES cell potency, and may thus provide a scientific basis for the ongoing debate surrounding the nuclear transfer technology. However, because the *Cdx2*-deficient embryo is not obviously abnormal before the onset of *Cdx2* expression, this approach may not solve the ethical dilemma. Moreover, research with primate or human cells will be required to assess whether *CDX2* is an optimal target for human application. Finally, we want to emphasize that ANT is a modification but not an alternative to nuclear transfer, as the approach requires additional manipulations of the donor cells that may complicate the logistics of production and safety assessment of patient-specific ES cell lines for therapy.

METHODS

Cloning and design of shRNAs. shRNAs were designed using the pSico-Oligomaker 1.5 (developed by A. Ventura, Jacks Lab), which is freely available at <http://web.mit.edu/ccr/labs/jacks/protocols/pSico.html>. Cloning into pSicoR was done as described on the website.

Generation of lentivirus and Cre-mediated recombination. Lentivirus was generated as described before⁶. The number of integrations was determined by Southern blot analysis. Genomic DNA was digested with *XbaI* (single cut in the viral backbone) and probed with an EGFP probe.

Cre-mediated recombination was achieved by transiently transfecting a Cre-recombinase-containing plasmid. Briefly, after introducing the Cre plasmid into the ES cells by electroporation, cells were cultured for 24 h (not longer, to avoid random integration of plasmid) in ES medium plus puromycin. GFP-negative subclones were picked, expanded and tested for recombination (*Cdx2*^{1Lox}).

Immunohistochemistry and RT-PCR. *Cdx2* staining of blastocysts was done as described previously⁵. The protocol is available on the Rossant laboratory website (<http://www.mshri.on.ca/rossant/protocols/immunoStain.html>). Monoclonal anti-*Cdx2* (CDX2-88, BioGenex) was used for all *Cdx2* stainings. RT-PCR was done with a one-step RT-PCR kit (Qiagen) using the following primers: β -actin 5'-GGCCCAGAGCAAGAGAGGTATCC-3' (forward) and 5'-ACGCACGATTTCCCTCTCAGC-3' (reverse); *Oct-4* (333 bp) 5'-GGATGGCA TACTGTGGACCT-3' (forward) and 5'-AGATGGTGGTCTGGCTGAAC-3' (reverse); and *Cdx2* (225 bp) 5'-AAACCTGTGCGAGTGGATG-3' (forward) and 5'-CTGCGGTTCTGAAACCAAT-3' (reverse). β -actin reverse transcription primer was published by ref. 5, and *Oct-4* and *Cdx2* primers were designed using PRIMER3.

NT, embryo transfer, ES cell derivation and 2N/4N blastocyst injections. Nuclear transfer was done as described previously^{8,17}. Nuclear transfer embryos were transferred at day 3.5 (morula/blastocyst stage) into the uteri of day 2.5 pseudo-pregnant recipient females. For ES cell derivation, the zona pellucida was removed using acidic tyrode (AT) solution and blastocysts were explanted on irradiated feeders in ES medium plus MEK1 inhibitor (PD98059). Diploid and tetraploid blastocyst injections were done as described previously¹⁸.

Received 23 August; accepted 3 October 2005.

Published online 16 October 2005.

1. Rideout, W. M. III, Hochedlinger, K., Kyba, M., Daley, G. Q. & Jaenisch, R. Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy. *Cell* **109**, 17–27 (2002).

2. Hwang, W. S. et al. Patient-specific embryonic stem cells derived from human SCNT blastocysts. *Science* **308**, 1777–1783 (2005).
3. Hurlbut, W. B. Altered nuclear transfer as a morally acceptable means for the procurement of human embryonic stem cells. *Perspect. Biol. Med.* **48**, 211–228 (2005).
4. Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J. & Beck, F. *Cdx2* is essential for axial elongation in mouse development. *Proc. Natl Acad. Sci. USA* **101**, 7641–7645 (2004).
5. Strumpf, D. et al. *Cdx2* is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093–2102 (2005).
6. Ventura, A. et al. Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl Acad. Sci. USA* **101**, 10380–10385 (2004).
7. Chawengsaksophak, K., James, R., Hammond, V. E., Kontgen, F. & Beck, F. Homeosis and intestinal tumours in *Cdx2* mutant mice. *Nature* **386**, 84–87 (1997).
8. Eggan, K. et al. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl Acad. Sci. USA* **98**, 6209–6214 (2001).
9. The President's Council on Bioethics. Alternative sources of human pluripotent stem cells. [Online] (<http://www.bioethics.gov>) (2005).
10. Melton, D. A., Daley, G. Q. & Jennings, C. G. Altered nuclear transfer in stem-cell research—a flawed proposal. *N. Engl. J. Med.* **351**, 2791–2792 (2004).
11. Jaenisch, R. Human cloning—the science and ethics of nuclear transplantation. *N. Engl. J. Med.* **351**, 2787–2791 (2004).
12. Brambrink, T., Hochedlinger, K. & Jaenisch, R. Gene expression in embryonic stem cells from cloned and fertilized embryos. *Proc. Natl Acad. Sci. USA* (submitted).
13. Adjaye, J. et al. Primary differentiation in the human blastocyst: Comparative molecular portraits of inner cell mass and trophectoderm cells. *Stem Cells* published online 4 August 2005 (doi:10.1634/stemcells.2005-0113).
14. Hyslop, L. A. et al. Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. *Stem Cells* **23**, 1035–1043 (2005).
15. Pfeifer, A., Ikawa, M., Dayn, Y. & Verma, I. M. Transgenesis by lentiviral vectors: lack of gene silencing in mammalian embryonic stem cells and preimplantation embryos. *Proc. Natl Acad. Sci. USA* **99**, 2140–2145 (2002).
16. Hacein-Bey-Abina, S. et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415–419 (2003).
17. Wakayama, S. et al. Establishment of male and female nuclear transfer embryonic stem cell lines from different mouse strains and tissues. *Biol. Reprod.* **72**, 932–936 (2005).
18. Wang, Z. & Jaenisch, R. At most three ES cells contribute to the somatic lineages of chimeric mice and of mice produced by ES-tetraploid complementation. *Dev. Biol.* **275**, 192–201 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank K. Hochedlinger and L. Jackson-Grusby for discussion and critical reading of the manuscript, and in particular D. Fu for sections and stainings. We are also grateful to B. Hogan and P. Berg for critical comments on the manuscript. R.J. was supported by NIH/NCI grants. A.M. was supported by a Boehringer Ingelheim Fonds (BIF) PhD fellowship.

Author Contributions R.J. and A.M. conceived and designed the experiments, A.M. performed the experiments, R.J. and A.M. wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.J. (jaenisch@wi.mit.edu).

LETTERS

Embryonic and extraembryonic stem cell lines derived from single mouse blastomeres

Young Chung^{1*}, Irina Klimanskaya^{1*}, Sandy Becker^{1*}, Joel Marh¹, Shi-Jiang Lu¹, Julie Johnson², Lorraine Meisner^{2,3} & Robert Lanza^{1,4}

The most basic objection to human embryonic stem (ES) cell research is rooted in the fact that ES cell derivation deprives embryos of any further potential to develop into a complete human being^{1,2}. ES cell lines are conventionally isolated from the inner cell mass of blastocysts^{3–5} and, in a few instances, from cleavage stage embryos^{6–9}. So far, there have been no reports in the literature of stem cell lines derived using an approach that does not require embryo destruction. Here we report an alternative method of establishing ES cell lines—using a technique of single-cell embryo biopsy similar to that used in pre-implantation genetic diagnosis of genetic defects¹⁰—that does not interfere with the developmental potential of embryos. Five putative ES and seven trophoblast stem (TS) cell lines were produced from single blastomeres, which maintained normal karyotype and markers of pluripotency or TS cells for up to more than 50 passages. The ES cells differentiated into derivatives of all three germ layers *in vitro* and in teratomas, and showed germ line transmission. Single-blastomere-biopsied embryos developed to term without a reduction in their developmental capacity. The ability to generate human ES cells without the destruction of *ex utero* embryos would reduce or eliminate the ethical concerns of many.

A series of six separate experiments was carried out to determine whether stem cell lines can be generated from single blastomeres (Supplementary Table 1). Eight-cell stage 129/Sv-ROSA26:*lacZ* mouse embryos were biopsied through a hole made in the zona pellucida with piezo-pulse drilling, and the biopsied (7-cell) embryos transferred to the oviducts of 1.5 days post coitum (d.p.c.) synchronized surrogate mothers. Each isolated blastomere was aggregated with a small clump of green fluorescent protein (GFP)-positive 129Sv/CD-1 mouse ES (mES) cells, and after incubation for 24–48 h, a growing 'bud' of GFP-negative cells was observed on the sides of the majority (60%) of GFP-mES clusters (Fig. 1a, b). The cell aggregates were plated onto mitomycin C-treated mouse embryonic fibroblasts (MEFs) and cultured in mES cell growth medium¹¹. Approximately half (36 out of 75) of them formed rapidly growing clumps of cells within 2–4 days, which were separated from GFP-positive mES cells by hand under a fluorescence microscope. The cells were expanded using mechanical and enzymatic methods, while further selecting by eye for the colonies morphologically resembling ES cells and excluding any GFP-positive cells (Fig. 1c–f). In four experiments, putative lines of LacZ⁺ ES cells were produced (Fig. 2a, c) that exhibited normal karyotype (Fig. 3g) and maintained markers of pluripotency after up to >50 passages. Each line expresses octamer binding protein 4 (Oct-4), stage-specific embryonic antigen (SSEA)-1, nanog and alkaline phosphatase (Fig. 2e, g, i).

Polymerase chain reaction (PCR) analysis confirmed the presence of LacZ but not GFP gene sequences in these cells (Supplementary Fig. 1a, b). Together with the karyotype analyses, the absence of GFP also rules out the possibility of contamination and/or fusion of the blastomere-derived lines with the ES cells used for co-culture.

When the putative ES cell cultures were allowed to overgrow or form embryoid bodies, they readily differentiated into cells of all three germ layers, as evidenced by immunostaining with antibodies to muscle actin (mesoderm, Fig. 3a), α -feto protein (primitive endoderm, Fig. 3c) and β III tubulin (ectoderm, Fig. 3e). Beating heart muscle, extraembryonic endoderm and multiple neuronal cell types were also routinely observed in differentiating cultures. To demonstrate further the pluripotency of the derived putative ES cells, cells were either injected into CD-1 mouse blastocysts or aggregated with 8-cell stage morulae as described¹¹. Forty-eight injected/aggregated embryos (between 8 and 15 per cell line) were transferred to recipient females. 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) staining of the resulting 29 fetuses (followed from mid-gestation through to term) showed that the ES cell lines contributed to all organ systems, including heart, kidney, liver, lung, intestine, brain, blood, skin and genital ridge, among others. Twenty-four of the fetuses (83%) were chimaeric (Fig. 3d, f), and eight out of nine (89%) pups (Fig. 3h) were chimaeric; the latter had the LacZ gene in their gametes (confirmed by PCR analysis; Supplementary Fig. 3), and produced LacZ⁺ offspring when crossed with CD-1 females, confirming the contribution of the blastomere-derived ES cells to the germ line. When the ES cells were injected into NOD-SCID (non-obese diabetic-severe combined immunodeficiency) mice, they formed teratomas containing tissues from all three germ layers, including bone and cartilage (mesoderm), neural rosettes (ectoderm), and ciliated respiratory epithelia (endoderm), among others (Fig. 3b).

Although stable putative ES stem lines were generated in only four of the six experiments, numerous other blastomere-derived outgrowths contained cells with both embryonic and extraembryonic stem-cell-like morphology. When FGF-4 was added to the media, seven putative TS lines were established, which maintained normal karyotype and expressed markers of TS cells (Fig. 2b, d, f, h, j). These cells were negative for Oct-4 (Fig. 2h) and for α -feto protein. RT-PCR analysis confirmed that these cells expressed *Cdx2*, but not Oct-4; *nanog* and *Rex-1* were expressed in both the putative TS and ES cell lines (data not shown). Putative TS cells contributed to the extraembryonic lineage in chimaeric fetuses generated by aggregation with the LacZ⁺ TS cells (Supplementary Fig. 2).

In two control experiments, individual blastomeres ($n = 44$) isolated from 8-cell embryos were plated into 20–100- μ l drops

¹Advanced Cell Technology, Worcester, Massachusetts 01605, USA. ²University of Wisconsin, State Laboratory of Hygiene, Molecular Cytogenetics, Madison, Wisconsin 53706, USA. ³Department of Pathology and Laboratory Medicine, University of Wisconsin Medical School, Madison, Wisconsin 53706, USA. ⁴Institute for Regenerative Medicine, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA.

*These authors contributed equally to this work.

containing mES cell culture medium. Most of the blastomeres failed to divide over the 10-day period of culture, whereas 9 (20%) generated small clusters of differentiated trophoblast-like giant cells (Fig. 1g, h) before arresting at the 2–6-cell stage, thus suggesting that cell co-culture was critical to the success of this system.

The blastomere-biopsied embryos in the present study developed to term without a reduction in their developmental capacity (49% (23 out of 47) live young versus 51% (38 out of 75) for control non-biopsied embryos (χ^2 test, $P = 0.85$)). These results are consistent with human data, which indicate that normal and pre-implantation genetic diagnosis (PGD)-biopsied embryos develop into blastocysts with comparable efficiency¹².

Although only 36 out of 125 blastomeres (29%) generated inner cell mass-like outgrowths, and only 12 stable putative ES and TS cell

lines were generated in this study (compared to approximately 25–35% for normal mouse blastocysts), we believe that this success rate can be considerably increased by greater attention to the earliest stages of blastomere outgrowth, as well as the use of various measures that influence the spontaneous differentiation of pluripotent ES cells into trophectoderm and other cell types. Owing to the labour intensiveness of the various steps, only conspicuous outgrowths were selected for further passaging, although many of the GFP-positive ES clusters were observed to contain other blastomere-derived stem cells. Preliminary experiments using all blastomeres at the 8-cell stage indicate that most of the blastomeres give outgrowths

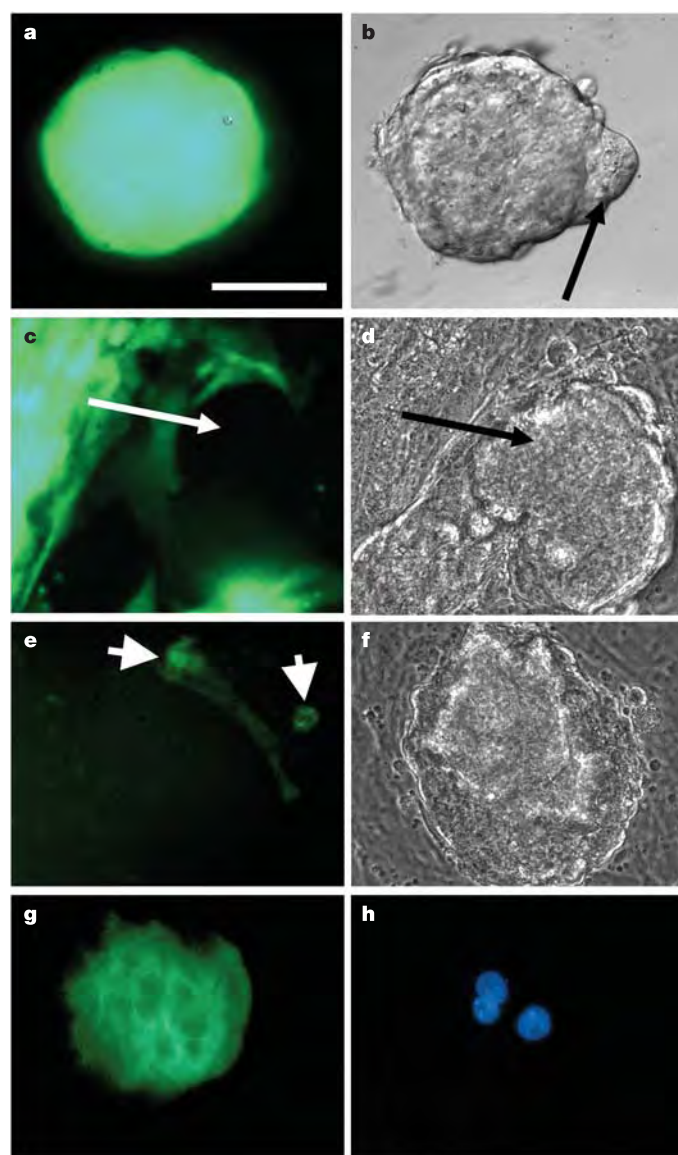


Figure 1 | Stages of single blastomere growth in the presence or absence of mES cells. **a, b**, Clump of GFP-positive mES cells 48 h after aggregation with single blastomeres; arrow shows a protruding cluster of GFP-negative cells. **c, d**, Outgrowth of GFP-negative cells aggregated with GFP-positive mES cells, after being plated on MEFs; arrows point to GFP-negative cells. **e, f**, Passage 1 of the outgrowth; arrows show remaining GFP-positive mES cells. **g, h**, Single blastomere outgrowth on MEFs for 4 days without ES cells, stained with Troma-1 and DAPI. **a, c, e, g**, Green fluorescence; **b, d, f, h**, Hoffman modulation optics; **d, f**, phase contrast. Scale bar, 100 μ m.

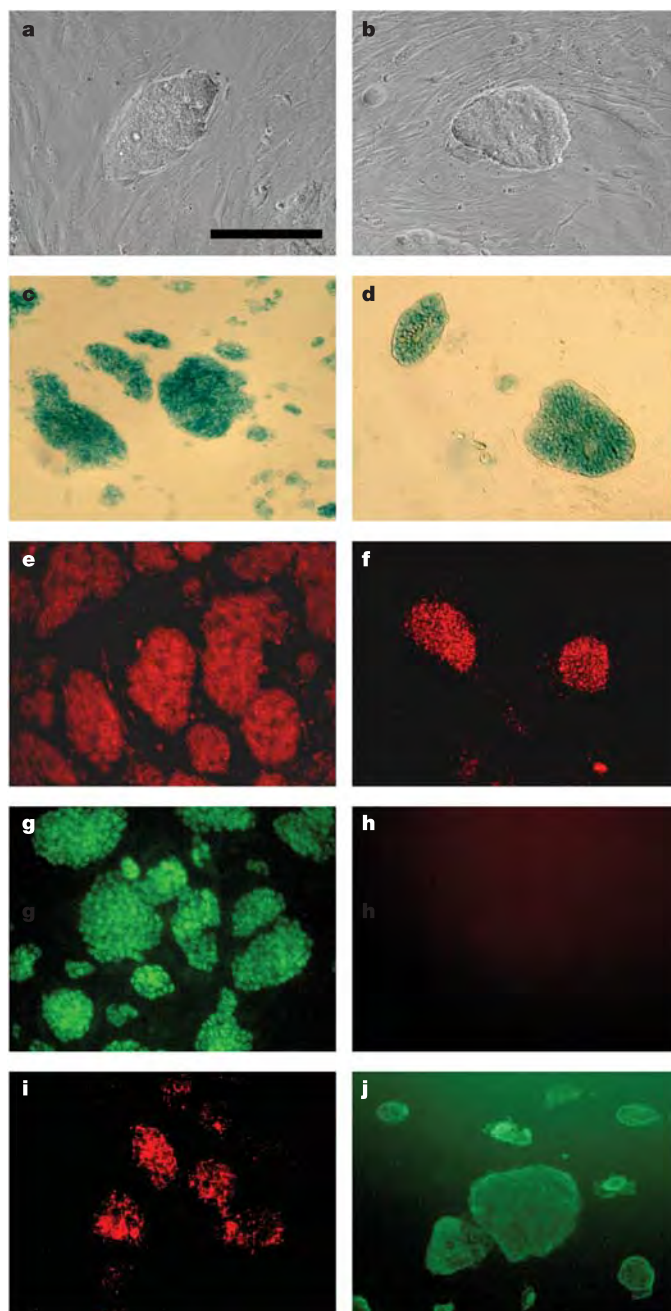


Figure 2 | Comparison of putative ES (left column) and TS (right column) cell lines derived from single blastomeres. **a, b**, Phase contrast photograph of typical colonies. **c, d**, LacZ-stained colonies, showing their single blastomere origin. **e, f**, Alkaline phosphatase staining. **g, h**, Indirect immunofluorescence with antibodies to Oct-4. **i**, SSEA-1 staining of putative ES cells; **j**, Troma-1 staining of putative TS cells (same field as **h**). Scale bar, 200 μ m.

that appear to be morphologically identical. However, it is unclear whether the individual blastomeres have different developmental fates based on these observations. The microenvironment is clearly critical, and variables such as ES cell number, the spatial orientation of blastomeres, ES cells and MEFs, as well as media composition, seem to be important.

Previous studies have described different approaches to establish pluripotent stem cell lines from blastocysts^{3–5} and 8–20-cell stage morulae^{6–9,13,14}. In two studies, stem cell lines were established from disaggregated morulae, although the cells from the embryos were cultured together and free to reaggregate^{9,13}. However, none of these strategies uses single isolated cells that could be used in conjunction with assisted reproduction technologies such as PGD, and, importantly, that would not interfere with the embryo's normal development to birth. There have been previous attempts in both humans and animals to induce single isolated blastomeres to proliferate *in vitro*, although none of these investigators derived pluripotent stem cell lines. One study¹⁵ cultured individual biopsied

mouse blastomeres *in vitro* on different extracellular matrix components, including fibronectin and laminin, which have been shown to promote attachment and proliferation of blastomeres *in vitro* and to enhance blastocyst development in various animal species^{16–18}. Most of the isolated blastomeres divided to form small sheets of 6–8 cells with a trophoblastic morphology similar to that described in ref. 19. In the human, another study²⁰ observed proliferation of blastomeres that had been removed and co-cultured with the biopsied embryos (range 1–8 cells per blastomere), although in all cases differential labelling indicated that they had generated trophoblastic cells. Unfortunately, the molecular mechanisms regulating these events are poorly understood, and it is unclear whether the success of the ES co-culture system in the present study is attributable to substances secreted by the ES cells or if cell–cell contact is required.

The developmental capacity of blastomeres isolated from mammalian embryos has been studied extensively, and it is clear that they retain their pluripotency, and, indeed, are capable of regular *in vivo* development upon transfer into mice²¹, rabbits²², sheep²³, swine^{18,24,25} and primates²⁶. We have demonstrated that single pre-implantation blastomeres can also be used to establish embryonic and extra-embryonic stem cell lines using an approach that does not interfere with the developmental potential of the parent embryo. The biopsy procedure described here is carried out in IVF clinics worldwide without a reduction in pregnancy rate. However, further investigations are required to determine whether stem cell lines can be derived from other mammalian species, including humans, using this micromanipulation technique. The ability to generate human ES cells from PGD blastomeres could circumvent the ethical concerns voiced by many, and allow the banking of autologous ES cell lines for children born from transferred embryos.

METHODS

Generation of ES and TS cell lines. Eight-cell stage 129/Sv-ROSA26:lacZ mouse embryos were biopsied through a hole in the zona pellucida using piezo-pulse drilling. The biopsied embryos were transferred to the oviducts of 1.5 d.p.c. synchronized surrogates, and each separated blastomere was aggregated with a small clump (≤ 100 cells) of GFP-positive 129Sv/CD-1 mES cells in a 300- μ m depression created by pressing a needle into the bottom of a plastic tissue culture plate, as described¹¹. After incubation for 24–48 h in mES cell growth medium¹¹ supplemented with 2,000 U ml⁻¹ mouse leukaemia inhibitory factor (LIF; Chemicon) and 50 μ M MEK1 inhibitor (Cell Signaling Technology), a growing bud of GFP-negative cells was observed on the sides of the majority of GFP-mES clusters. The aggregates were plated onto mitomycin C-treated MEFs and cultured in mES cell growth medium until GFP-negative clumps became large enough for dispersion (> 20 cells), which then were separated from GFP-positive mES cells by hand with a microcapillary under a fluorescence microscope. The cells were dissociated and expanded by alternating mechanical dissociation and digestion with 0.05% trypsin (Invitrogen). Blastomere outgrowths that morphologically resembled trophoblast and extraembryonic endoderm but not ES cells were further cultured in the mES cell medium with 50 ng ml⁻¹ FGF-4, and produced TS-like cells that were maintained under these conditions and passaged with trypsin.

Immunofluorescence and alkaline phosphatase staining. Indirect immunofluorescence staining was performed on cells growing on 4-well tissue culture plates as previously described^{2,5}. The following primary antibodies were used: Oct-4 (Santa Cruz Biotechnology), SSEA-1 (developed by D. Solter and B. Knowles and obtained through the DSHB (Developmental Studies Hybridoma Bank) of the University of Iowa), Troma-1 (raised by P. Brulet and R. Kemler and obtained through DSHB), α -feto protein (DACO), β III tubulin (Covance) and muscle actin (Abcam). Alkaline phosphatase staining was performed using the Vector red kit from Vector Laboratories.

Chimaeras and X-gal staining. Approximately 15 ES cells or TS cells were injected either into the blastocoels of expanded blastocysts or into the perivitelline space of pre-compacting 8-cell stage embryos obtained from CD-1 mice. Blastocyst chimaeras were transferred into the uteri of 2.5 d.p.c. surrogate mice 4 h after the ES cell injections. Most of the 8-cell stage chimaeric embryos developed into expanded blastocysts after overnight culture in mES medium, and were then transferred to surrogate mice. The pregnancies were terminated at days 11.5, 12.5 or 19.5, and the embryos and placentas were fixed overnight in 4%

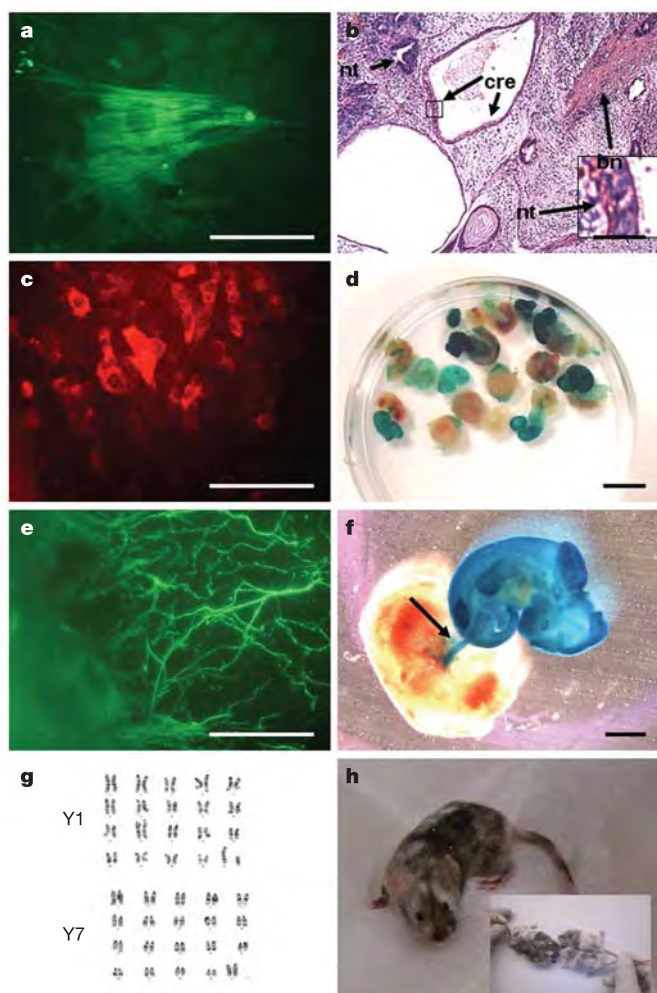


Figure 3 | Differentiation of blastomere-derived mES cells *in vitro* and *in vivo*. **a, c, e,** Immunofluorescence analysis of molecular markers of mesoderm (muscle actin, **a**), primitive endoderm (α -feto protein, **c**) and ectoderm (β III tubulin, **e**). **g,** Representative chromosome spreads of two single-blastomere-derived mES cell lines. **b,** Teratomas section, stained with haematoxylin and eosin. bn, bone (mesoderm); nt, neural tissue (ectoderm); cre and insert, ciliated respiratory epithelium (endoderm). **d, f, h,** 11.5 d.p.c. chimaeric embryos (**d, f**) and chimaeric pups (**h**) produced from three mES cell lines; arrow in **f** shows the extraembryonic mesoderm-derived placental labyrinth, also chimaeric. Scale bars: **a, c,** 100 μ m; **e,** 200 μ m; **d,** 10 mm; **f, h,** 2 mm.

paraformaldehyde and washed with PBS overnight; LacZ staining using a X-gal staining kit (GAL-S, Sigma) was done to test for chimaerism in the embryos.

Teratomas. Approximately 1 million ES cells were injected into the rear thigh of a NOD-SCID mouse. After approximately 2 months the mice were killed and the teratomas excised, fixed in 4% paraformaldehyde, embedded in paraffin and sectioned.

Received 18 August; accepted 4 October 2005.

Published online 16 October 2005.

- Committee on the Biological and Biomedical Applications of Stem Cell Research, Board on Life Sciences, National Research Council, Board on Neuroscience and Behavioral Health, Institute of Medicine (ed.) *Stem Cells and the Future of Regenerative Medicine* (National Academies Press, Washington DC, 2002).
- Lanza, R. et al. (eds) *Handbook of Stem Cells. Vol 1: Embryonic Stem Cells* (Elsevier/Academic Press, San Diego, California, 2004).
- Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
- Thomson, J. A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
- Cowan, C. A. et al. Derivation of embryonic stem-cell lines from human blastocysts. *N. Engl. J. Med.* **350**, 1353–1356 (2004).
- Strelchenko, N. et al. Morula-derived human embryonic stem cells. *Reprod. BioMed.* **9**, 623–629 (2004).
- Sukoyan, M. A. et al. Embryonic stem cells derived from morulae, inner cell mass, and blastocysts of mink: comparisons of their pluripotencies. *Mol. Reprod. Dev.* **36**, 148–158 (1993).
- Mitalipova, M., Beyhan, Z. & First, N. L. Pluripotency of bovine embryonic cell line derived from precompacting embryos. *Cloning* **3**, 59–67 (2001).
- Delhaise, F. et al. Establishment of an embryonic stem cell line from 8-cell stage mouse embryos. *Eur. J. Morphol.* **34**, 237–243 (1996).
- Handyside, A. H. et al. Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature* **344**, 768–770 (1990).
- Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. (eds) *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002).
- Staessen, C. et al. Comparison of blastocyst transfer with or without preimplantation genetic diagnosis for aneuploidy screening in couples with advanced maternal age: a prospective randomized controlled trial. *Hum. Reprod.* **19**, 2849–2858 (2004).
- Eisetter, H. R. Pluripotent embryonic stem cell lines can be established from disaggregated mouse morulae. *Dev. Growth Differ.* **31**, 275–282 (1989).
- Tesar, P. J. Derivation of germ-line-competent embryonic stem cell lines from preblastocyst mouse embryos. *Proc. Natl Acad. Sci. USA* **102**, 8239–8244 (2005).
- Wilton, L. & Trounson, A. Biopsy of preimplantation mouse embryos: development of micromanipulated embryos and proliferation of single blastomeres *in vitro*. *Biol. Reprod.* **40**, 145–152 (1989).
- Tao, T. & Niemann, H. Cellular characterization of blastocysts derived from rabbit 4-, 8- and 16-cell embryos and isolated blastomeres cultured *in vitro*. *Hum. Reprod.* **15**, 881–889 (2000).
- Armant, D. R., Kaplan, H. A. & Lernatz, W. J. Fibronectin and laminin promote *in vitro* attachment and outgrowth of mouse blastocysts. *Dev. Biol.* **116**, 519–523 (1986).
- Saito, S. & Niemann, H. Effects of extracellular matrices and growth factors on the development of isolated porcine blastomeres. *Biol. Reprod.* **44**, 927–936 (1991).
- Tarkowski, A. K. & Wroblewska, J. Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. *J. Embryol. Exp. Morphol.* **18**, 155–180 (1967).
- Geber, S., Winston, R. M. L. & Handside, A. H. Proliferation of blastomeres from cleavage stage human embryos *in vitro*: an alternative to blastocyst biopsy for preimplantation diagnosis. *Hum. Reprod.* **10**, 1492–1496 (1995).
- Rossant, J. Postimplantation development of blastomeres isolated from 4- and 8-cell mouse eggs. *J. Embryol. Exp. Morphol.* **36**, 283–290 (1976).
- Moore, N. W., Adams, C. E. & Rowson, L. E. Developmental potential of single blastomeres of the rabbit egg. *J. Reprod. Fertil.* **17**, 527–531 (1968).
- Willadsen, S. M. The development capacity of blastomeres from 4- and 8-cell sheep embryos. *J. Embryol. Exp. Morphol.* **15**, 165–172 (1981).
- Menino, A. R. & Wright, R. W. Effect of pronase treatment, microdissection, and zona pellucida removal on the development of porcine embryos and blastomeres *in vitro*. *Biol. Reprod.* **28**, 433–446 (1983).
- Niemann, H. & Reichelt, B. Manipulating early pig embryos. *J. Reprod. Fertil. Suppl.* **48**, 75–94 (1993).
- Chan, A. W. S. et al. Clonal propagation of primate offspring by embryo splitting. *Science* **287**, 317–319 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The monoclonal antibodies to SSEA-1 (developed by D. Solter and B. Knowles) and to Troma1 (developed by P. Bruet and R. Kemler) were obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com. Correspondence and requests for materials should be addressed to R.L. (rlanza@advancedcell.com).

LETTERS

Planar cell polarity signalling couples cell division and morphogenesis during neurulation

Brian Ciruna^{1†}, Andreas Jenny², Diana Lee¹, Marek Mlodzik² & Alexander F. Schier^{1†}

Environmental and genetic aberrations lead to neural tube closure defects (NTDs) in 1 out of every 1,000 births¹. Mouse and frog models for these birth defects have indicated that Vangl2 (Vangl2, also known as Strabismus) and other components of planar cell polarity (PCP) signalling might control neurulation by promoting the convergence of neural progenitors to the midline^{2–8}. Here we show a novel role for PCP signalling during neurulation in zebrafish. We demonstrate that non-canonical Wnt/PCP signalling polarizes neural progenitors along the anteroposterior axis. This polarity is transiently lost during cell division in the neural keel but is re-established as daughter cells reintegrate into the neuroepithelium. Loss of zebrafish Vangl2 (in *trilobite* mutants) abolishes the polarization of neural keel cells, disrupts re-intercalation of daughter cells into the neuroepithelium, and results in ectopic neural progenitor accumulations and NTDs. Remarkably, blocking cell division leads to rescue of *trilobite* neural tube morphogenesis despite persistent defects in convergence and extension. These results reveal a function for PCP signalling in coupling cell division and morphogenesis at neurulation and indicate a previously unrecognized mechanism that might underlie NTDs.

During zebrafish neurulation the neural plate folds towards the midline. This results in the apposition of apical surfaces from opposite sides of the neural plate and the formation of the neural keel (Supplementary Fig. 1). As cells divide, one daughter cell remains in the ipsilateral side of the neural keel, whereas the other daughter cell intercalates across the midline and integrates into the contralateral neuroepithelial layer^{9–11}. To explore the molecular basis of neural progenitor cell morphogenesis we used a candidate gene approach and examined whether the PCP signalling component Vangl2 might be involved^{12,13}. We eliminated all Vangl2 activity by generating maternal-zygotic *trilobite* (MZtri) mutants with the use of a germline-replacement strategy¹⁴. MZtri embryos proved more severely affected than zygotic mutants (Supplementary Fig. 2). Comparison of wild-type (WT) and mutant embryos at the 20-somite stage revealed that MZtri embryos do not generate a normal neural tube (Fig. 1g, h). The MZtri neural anlage develops as an outer pseudo-stratified neuroepithelial layer surrounding an ectopic mass of disorganized cells (Fig. 1h). As early as the neural keel stage, the MZtri neural primordium appears broader and thicker than in WT (Fig. 1c, d). This trend continues through neural rod stages when cells seem to accumulate in the centre of the wide MZtri neural anlage (Fig. 1e, f). The floorplate of MZtri mutant embryos also appears broader than in WT (Fig. 1e–h), as is evident in sections through *sonic hedgehog*-stained WT and MZtri embryos (Supplementary Fig. 8e, g). Expanded neural midline structures are also characteristic of frog and mouse PCP signalling mutants^{2,8}.

Because Vangl2 has been shown to modulate the non-canonical Wnt signalling pathway, we asked whether Wnt signals also regulate neural tube morphogenesis. Using a modified germline-replacement protocol (Supplementary Fig. 3), we generated *wnt11/silberblick* (*slb*)¹⁵ and *wnt5/pipetail* (*ppt*)¹⁶ MZ compound mutants (Supplementary Fig. 4). MZslb/MZppt embryos showed a similar, yet less severe, neurulation phenotype to that of MZtri mutants (Fig. 1i). Reduction of Wnt4 activity in an MZslb/MZppt background (through injection of *wnt4* antisense morpholino oligonucleotides¹⁷) enhanced the mutant phenotype, and at the 20-somite stage MZslb/MZppt;*wnt4*-morphant embryos displayed a neurulation phenotype very similar to that of MZtri mutants (Fig. 1j, and Supplementary Fig. 4). These results indicate that non-canonical Wnt signalling is required for normal zebrafish neurulation.

In the frog, neural tube closure requires PCP signalling within the neural plate¹⁸. To determine whether MZtri neurulation defects are autonomous to the neuroectoderm or secondary to mesoderm or endoderm convergence and extension defects, we examined neurulation in embryos lacking endoderm and trunk and head mesoderm. Such embryos were generated by misexpression of Lefty, an inhibitor of Nodal signalling^{19,20}. MZtri + *lefty* embryos were considerably shorter than WT + *lefty* controls (compare Fig. 2a with Fig. 2b), and had neurulation defects similar to MZtri mutants (Fig. 2b, inset). In a complementary assay, we examined whether mutant mesoderm can induce the MZtri neurulation phenotype. We generated chimaeric embryos in which only the endoderm and trunk mesoderm lineages were derived from MZtri mutant cells (Fig. 2c). In these embryos the neural tube developed with normal neuroepithelial morphology, a well-formed neurocoel and no evidence of ectopic cell accumulations (Fig. 2c'). These results indicate that MZtri neurulation defects are due to the lack of Vangl2 function in ectodermal tissues.

Several potential mechanisms might underlie the ectopic accumulation of cells seen in MZtri mutants, including abnormal delamination of neuroepithelial cells or failed reintegration of cells into the neuroepithelium after cell division. As a first test to distinguish between these possibilities, we used the photoconvertible Kaede fluorophore²¹ to label half of the neuroepithelium at neural plate or early neural keel stages and then analysed the location of the labelled cells and their descendants in the neural tube (Fig. 3a–d). In agreement with previous studies of zebrafish neurulation^{9–11}, we observed that cell division in the neural keel results in the bilateral distribution of daughter cells across apposing neuroepithelial layers of the WT neural tube ($n = 10$; Fig. 3a, b). In marked contrast, labelled cells were not found in the contralateral neuroepithelium of MZtri embryos ($n = 29$), and a sharp midline boundary was maintained even among cells accumulating ectopically in the neural anlage

¹Developmental Genetics Program, Skirball Institute of Biomolecular Medicine and Department of Cell Biology, New York University School of Medicine, New York, New York 10016, USA. ²Mount Sinai School of Medicine, Brookdale Department of Molecular, Cellular and Developmental Biology, 1 Gustave L. Levy Place, New York, New York 10029, USA. †Present addresses: Program in Developmental Biology, The Hospital for Sick Children, Toronto Medical Discovery Tower, Toronto, Ontario M5G 1L7, Canada (B.C.); Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA (A.F.S.).

(Fig. 3c, d). These results are consistent with a defect in the integration of MZtri neural progenitors into the contralateral neuroepithelium.

To follow more directly the behaviour of neural progenitors during and after mitosis, we imaged cell behaviour in the neural keel of WT ($n = 14$) and MZtri mutant ($n = 11$) embryos (Fig. 3e–m). As observed previously, we found that WT cells rounded up and divided apically—that is, along the medial–lateral axis of the neural keel—and that daughter cells became incorporated into opposite sides of the neural tube (Fig. 3e–g, and Supplementary movie 1)^{10,11}. In addition, we found that the more basal daughter cell maintained contact with the basement membrane through a thin cellular process (Fig. 3e, f, arrow) and returned to its original position within the neuroepithelium. In contrast, the more apical daughter cell lost contact with the basement membrane, became polarized across the medial–lateral axis and intercalated across the midline into the contralateral side of the neural keel (Fig. 3f, g, arrowhead). Intercalation of apical daughter cells across the midline took an average of 8 min after completion of cytokinesis (data not shown).

In MZtri embryos, cell division seemed normal. Cells divided apically, and basal daughter cells maintained their ipsilateral position within the neuroepithelium (Fig. 3h, i, arrow). In notable contrast to

the wild type, however, apical MZtri daughter cells failed to reintegrate into the neuroepithelium after mitosis (Fig. 3i, j, arrowhead). These daughter cells accumulated in the middle of the MZtri neural anlage and remained in the place where they were formed (Fig. 3j, and Supplementary Fig. 5). To determine whether MZtri intercalation defects were secondary to abnormal neural tube morphogenesis, we examined cell division at the onset of neural keel formation, before an obvious MZtri neurulation phenotype (compare Fig. 1a, b). We observed that after mitosis, apical MZtri daughter cells failed to intercalate across the midline (Fig. 3k–m, and Supplementary movie 2). These results indicate that the PCP pathway is required for the intercalation of neural progenitor cells into the contralateral neuroepithelial layer after cell division in the neural keel.

PCP signalling functions to polarize cells, but molecular evidence for such a role during neurulation has been elusive. We therefore analysed the subcellular localization of PCP signalling components^{22,23} and examined whether the failure of MZtri cell reintegration is due to polarity defects. We found that enhanced green fluorescent protein (EGFP)-tagged Prickle²⁴ (Gfp-Pk), a PCP effector molecule, showed striking asymmetric localization in WT cells of the notochord and neural keel (Fig. 4a, d). This fusion protein was functional and rescued zebrafish Pk1-morphants

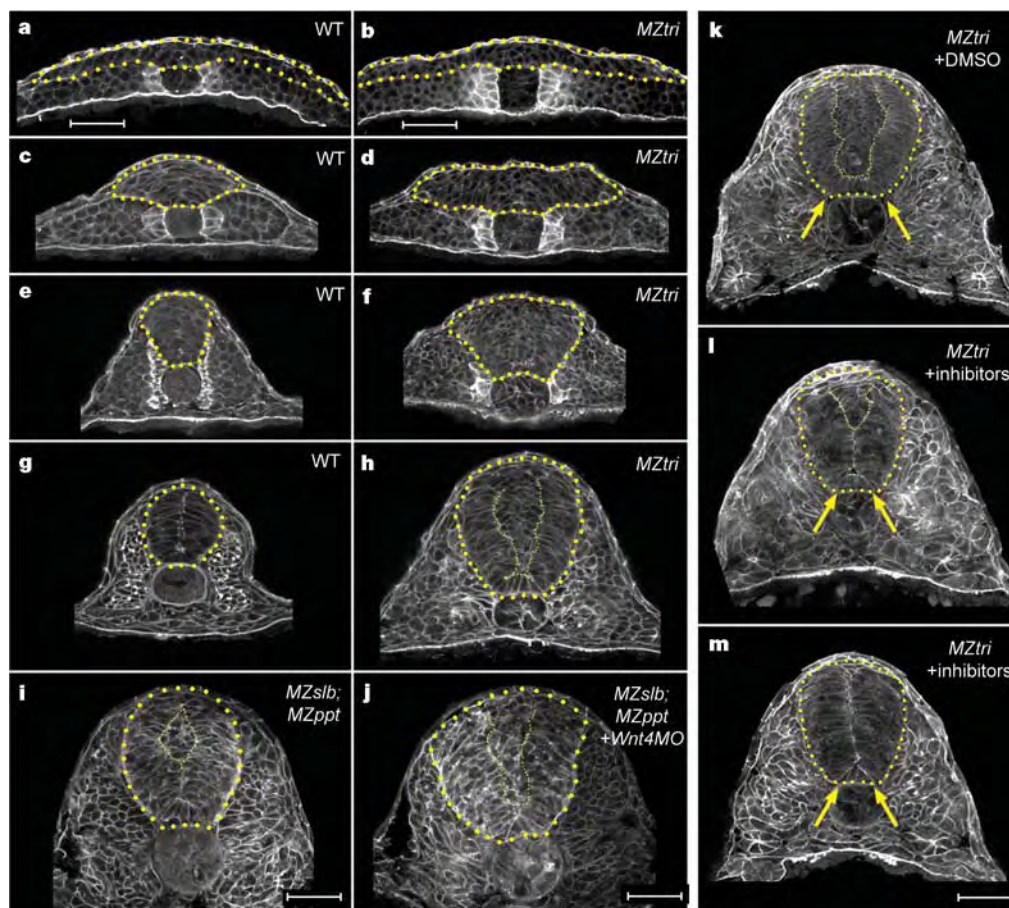


Figure 1 | PCP signalling is required for zebrafish neural tube formation. **a–h**, Confocal micrographs of transverse sections through rhodamine-phalloidin-stained embryos, comparing WT and MZtri neural tube morphogenesis at five-somite/neural plate (**a**, **b**), ten-somite/neural keel (**c**, **d**), 15-somite/neural rod (**e**, **f**) and 20-somite/neural tube (**g**, **h**) stages. The neural anlage is outlined in all images. **i**, **j**, Sections through mGFP-injected 20-somite-stage embryos showing ectopic cell accumulations within the developing neural tube of MZslb;MZppt embryos (**i**), and severe disorganization of the neural anlage in MZslb;MZppt

embryos that had been injected with 6 ng of Wnt4 morpholino antisense oligonucleotides (**j**). **k–m**, Rhodamine-phalloidin-stained sections through 20-somite-stage MZtri embryos cultured overnight either in 4% dimethylsulphoxide (**k**) or in the presence of the DNA synthesis inhibitors aphidicolin and hydroxyurea²⁵ (**l**, **m**). Note the rescue of the expanded floorplate (arrows in **k–m**) and defects in neural tube morphogenesis on blocking of cell division. The extent of ectopic cellular accumulations in PCP signalling mutants is outlined (**h–l**). Scale bars, 50 μ m.

(Supplementary Fig. 6). Gfp-Pk was present predominantly in the cytoplasm but became asymmetrically localized to distinct, dynamic puncta along the anterior membrane of neural keel cells (Fig. 4a, arrows, and Supplementary movie 3). Membrane localization of Gfp-Pk was lost during cell division in the neural keel (Fig. 4e) but was subsequently re-established in both daughter cells (Fig. 4f). To determine whether this localization is dependent on PCP signalling,

we analysed *MZtri* and *MZslb;MZppt;wnt4*-morphant embryos (Fig. 4b, c). In both contexts the asymmetric localization of Gfp-Pk was severely reduced or absent. These results establish membrane localization of Gfp-Pk as the first molecular marker of planar polarity in neural progenitors and reveal that PCP signalling polarizes cells across the anteroposterior (AP) axis.

To determine the cell autonomy of Vangl2 function and Gfp-Pk localization during neurulation, we generated chimaeras of WT and *MZtri* cells (Supplementary Table 1). When WT donor cells were transplanted into WT hosts, most donor clones distributed daughter cells into the contralateral side of the neural tube (88% of clones, $n = 16$; Fig. 2d). In contrast, only a minority of *MZtri* donor clones in the neural tube of WT hosts (38%, $n = 37$) showed contralateral cell intercalation, despite the normal morphogenesis of the surrounding WT neuroepithelial tissue (Fig. 2e). Time-lapse analysis of *MZtri* cell behaviour in WT neural keels showed that most *MZtri* daughter cells ($n = 4$ of 6) failed to intercalate across the midline after mitosis, indicating a cell-autonomous role for Vangl2 (Supplementary Fig. 7a–c). *MZtri* cells transplanted into WT host embryos also failed to localize Gfp-Pk to the membrane (Fig. 4g). In cases where *MZtri* cells showed some evidence of medial–lateral intercalation, the rate of cell movement was severely delayed ($n = 2$ of 6; Supplementary Fig. 7d–f).

Most *MZtri* donor clones in *MZtri* hosts showed no evidence of cell intercalation across the midline (95%, $n = 41$; Fig. 2f). Similarly, in transplantations of WT cells into *MZtri* hosts, 93% of donor clones did not intercalate into the contralateral neuroepithelium ($n = 61$; Fig. 2g). Time-lapse analysis of WT cell behaviour in a *MZtri* neural keel confirmed that WT neural progenitors fail to polarize and re-intercalate after mitosis ($n = 7$; Supplementary Fig. 7g–i). Strikingly, WT cells transplanted into *MZtri* hosts also showed a decrease in membrane-localized Gfp-PK and a loss in AP polarity (Fig. 4h), indicating that non-autonomous effects on neurulation are not simply the result of a passive obstruction to cell intercalation. These results indicate that Vangl2 might be required both autonomously in cells that must reintegrate into the neuroepithelium and non-autonomously in neighbouring cells in order to establish or maintain planar polarity in neural progenitors.

Our results show that PCP signalling is required for the repolarization and reintegration of neural progenitors after cell division in the neural keel. Indeed, the predominant role of PCP signalling might be to counteract the morphogenetic consequences of mitosis, which results in loss of polarity and the exclusion of apical daughter cells from the neuroepithelium. The strictest inference of this model would be that *MZtri* neurulation defects may be suppressed by blocking cell division, thus precluding the need for PCP signalling. We therefore attempted to block cell division in WT and *MZtri* embryos through application of the DNA synthesis inhibitors aphidicolin and hydroxyurea²⁵. Mitotic inhibitors were applied at late gastrulation stages to target maximal effects on neurulation. Treatment with inhibitor significantly decreased the incidence of contralateral cell intercalation of WT neural progenitors (Supplementary Table 1), but neurulation proceeded normally in treated WT embryos (Supplementary Fig. 8). Strikingly, blocking cell division suppressed *MZtri* neurulation defects in 90% of embryos ($n = 41$; Fig. 1k–m, and Supplementary Fig. 8) and rescued neural tube morphogenesis, aberrant floorplate expansion and ectopic neural progenitor accumulations in 62% of these cases (Fig. 1m). This result indicates that PCP signalling is no longer required for neural tube formation if cell division is blocked.

Taken together, our studies indicate that through re-establishing cell polarity and directing intercalative behaviour, PCP signalling might function to correct for the morphogenetic consequences of mitoses on neural tube morphogenesis. First, in the neural plate, midline progenitors distribute daughter cells along the medial–lateral axis, thus broadening the neural midline and future floor-

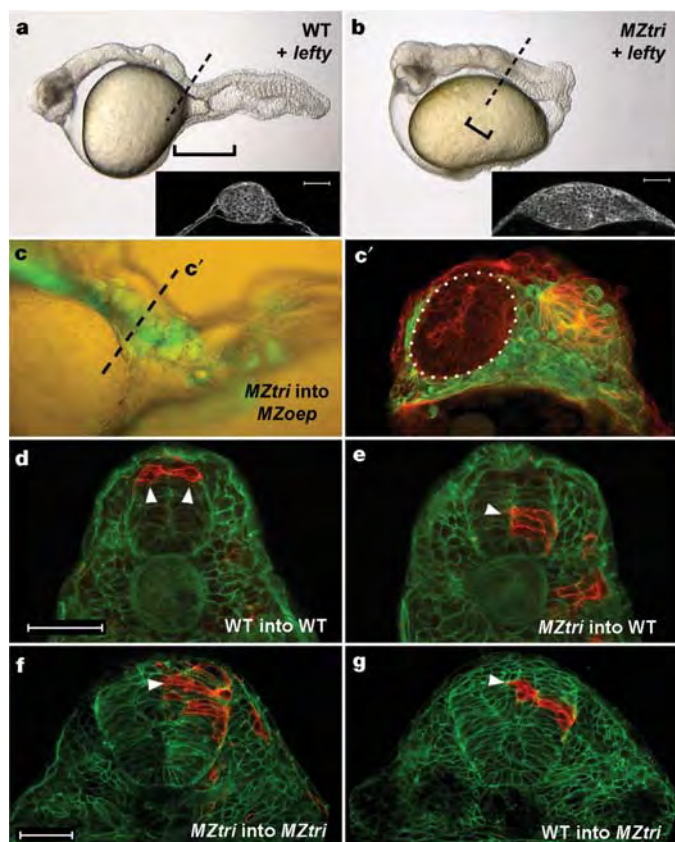


Figure 2 | Cell autonomy of PCP signalling within the neural keel.

a, b, Whole mounts and transverse sections through the trunk of WT (**a**) and *MZtri* (**b**) embryos 24 h after fertilization, injected with 100 pg of *lefty* mRNA. Convergence of the neural plate into a neural rod occurs normally in WT + *lefty* embryos (**a**), despite the absence of underlying mesendoderm (inset, section indicated by dotted line). This convergence is disrupted in *MZtri* + *lefty* mutants (**b**), which show neurulation defects in the absence of trunk mesoderm (inset, section indicated by dotted line). Brackets in **a** and **b** indicate the extent of trunk axial extension. Identical results were obtained with a different genetic combination: maternal-zygotic *one-eyed-pinhead* (*MZoe*) mutants were used to eliminate Nodal signalling, and PCP signalling was perturbed through the injection of *diego* mRNA (not shown). **c**, mGFP-labelled *MZtri* cells were transplanted into *MZoe* host embryos at mid-blastula stages to generate *MZtri* → *MZoe* chimaeric embryos. *MZoe* mutants lack Nodal signalling and do not form endoderm or trunk mesoderm lineages²⁹; endoderm and trunk somites in chimaeric embryos therefore develop entirely from GFP-positive *MZtri* donor cells. **c'**, Transverse sections of *MZtri* → *MZoe* chimaeras 24 h after fertilization were counterstained with rhodamine-phalloidin to visualize neural tube formation (outlined with a dotted line). The absence of *MZtri* neurulation defects indicates a requirement for PCP signalling autonomous to the neuroectoderm. **d–g**, Transverse sections through the trunk of WT and *MZtri* chimaeric embryos at 24 h after fertilization. **d**, Bilateral distribution of mRFP-labelled WT cells (arrowheads) within the neural tube of an mGFP-labelled WT host. **e**, Unilateral accumulation of mRFP-labelled *MZtri* cells within the neural tube of mGFP-labelled WT host embryos. **f, g**, mRFP-labelled *MZtri* (**f**) or WT (**g**) cells transplanted into mGFP-labelled *MZtri* host embryos accumulate unilaterally within the *MZtri* neural anlage. Scale bars, 50 μ m.

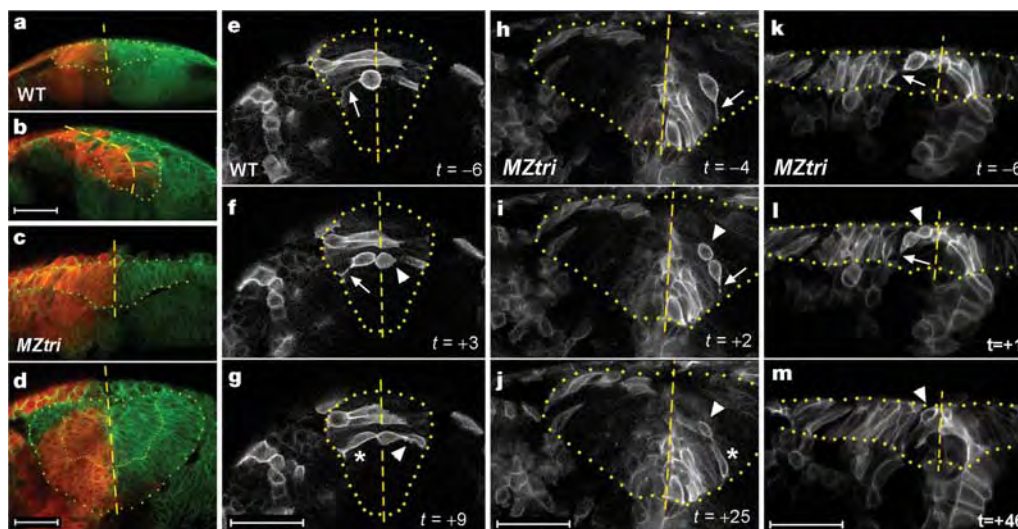


Figure 3 | The cellular basis of MZtri neurulation defects. **a–d**, Lineage tracing of WT (**a**, **b**) and MZtri (**c**, **d**) neural plate cells after red photoconversion of the Kaede fluorophore. WT neural progenitors routinely crossed the midline into the contralateral side of the neural tube (**b**). MZtri neural progenitor cells never integrated into the contralateral neuroepithelium (**d**). Cells accumulating ectopically in the centre of the MZtri neural anlage also respected the midline (dashed lines). **e–m**, Confocal micrographs from time series depicting cell division during WT (**e–g**) and MZtri (**h–m**) neurulation. Time *t* is indicated as minutes before or after the completion of cytokinesis. The boundary of the neural keel has been highlighted, and the midline indicated. Cells in the WT neural

keel round up their cell bodies and divide apically (**e**). Basal daughter cells remain connected to the basement membrane through a cellular process (arrow in **e** and **f**) and reinsert into the neuroepithelium (asterisk in **g**). Apical daughter cells (arrowhead in **f** and **g**) lose contact within the basal membrane, adopt medial–lateral polarity and intercalate across the midline of the neural keel. MZtri cells divide apically, and basal daughter cells behave as in WT (arrow in **h**, **i**, **k** and **l**; asterisk in **j**). MZtri apical daughter cells (arrowhead in **i**, **j**, **l** and **m**) do not intercalate into the contralateral neuroepithelium and remain in the place where they were formed. Scale bars, 50 μ m.

plate¹¹. PCP-mediated convergence of midline cells counters this broadening^{2–8}. Second, neural keel cells divide apically and position daughter cells outside the neuroepithelium. PCP-mediated reintegration of neural progenitors ensures that these cells do not accumulate in the midline. Although our results do not exclude additional roles for PCP signalling during neural development, blocking cell division abrogates the need for PCP signalling during neurulation and suppresses the floorplate and neural tube phenotypes associated with MZtri mutants.

Our results contrast with a recent study²⁶ that analysed the role of PCP signalling during earlier stages of zebrafish development. It was shown that PCP signalling functions upstream of mitosis to orient the plane of cell division at gastrulation. At this stage, PCP signals instruct neural precursors to divide along the animal–vegetal axis, thus driving extension of the zebrafish axis. Conversely, our study indicates that PCP signalling functions after mitosis to counteract the morphogenetic consequences of cell divisions during neurulation. Although the molecular basis of this process is unknown, it is possible that PCP signalling regulates molecules involved in the establishment of apical–basal polarity or cell adhesion, two processes that might be disrupted by mitosis²⁷. It also remains to be determined whether components of the mitotic apparatus directly regulate the re-establishment of planar polarity. No matter what the precise molecular underpinnings might be, our study shows that PCP signalling is required to compensate for the morphogenetic consequences of cell division on neural tube morphogenesis by promoting the polarity and intercalation of neural progenitors. Neural tube closure defects in *curly tail* (*ct*) mice, a genetic model for human NTDs, can also be modulated with agents that slow the rate of embryonic cell division²⁸. Pharmacological inhibition of cell division at gastrula stages exacerbates the *Ct* NTD phenotype, whereas treatment during neurulation rescues closure defects of the neural tube²⁸. Although the nature of the *Ct* mutation and the mechanism of pharmacological rescue remain unclear, our results raise the possibility that the uncoupling of PCP signalling and cell division during

neurulation represent a common origin of neural tube closure defects.

METHODS

Strains. The following mutants were used: *ppt*(*sk13*) (Supplementary Fig. 4), *slb*(*tx226*)¹⁵, *tri*(*tk50f*)¹³ and MZ*oe*(*tz57*)²⁹. Germline-replacement chimaeras for *tri* were generated as described previously¹⁴. The generation of germline-replacement chimaeras for *slb*; *ppt* compound mutants is described in Supplementary Fig. 3.

Embryo microinjection and transplantation. Plasmids containing membrane-localized GFP (*mGFP*), membrane-localized red fluorescent protein RFP (*mRFP*), EGFP-tagged Prickle (*Gfp-Pk*)²⁴, *Kaede*²¹ or *lefty*^{19,20} were linearized and sense-strand-capped mRNA was synthesized with the mMACHINE mMESSAGE system (Ambion). Morpholino antisense oligonucleotides (Gene Tools) were designed for zebrafish *Pk1* (ref. 30) and *Wnt4* (ref. 17) as described previously. Zebrafish embryos were dechorionated by treatment with pronase and injected at the one-cell stage. Scatter labelling was obtained by injecting a subset of blastomeres at the 16-cell to 32-cell stage. Cell transplantations were performed at mid-blastula stages, as described previously¹⁴. For chimaeric analyses of Vangl2 function, one-cell-stage WT and MZtri embryos were injected with 100 pg of either *mRFP* or *mGFP* mRNA. At mid-blastula stages, about 30–50 donor cells were transplanted into a single location above the margin of host embryos, to ensure unilateral distribution of donor cell clones within the anterior spinal cord of host embryos.

Cell division inhibitors. To block cell division at neurulation, embryos were cultured in a solution of 150 μ M aphidicolin (Sigma) and 20 mM hydroxyurea (Sigma) in 4% dimethylsulphoxide²⁵, beginning at 80% epiboly stages.

Sectioning and microscopy. For transverse sections, embryos were fixed overnight in 4% paraformaldehyde, embedded in 2% agarose and sectioned on a vibratome into 200- μ m slices. Live embryos were mounted in 0.8% agarose before imaging. Fluorescent images of embryos injected with mGFP, mRFP or Gfp-Pk, or samples stained with rhodamine-phalloidin (Molecular Probes) were obtained with a Zeiss LSM510 confocal microscope. For live imaging of cell divisions within the neural keel, embryos were scatter-labelled with mGFP and imaging was performed in a transverse plane through the trunk of 6–12-somite-stage embryos at the first to sixth somite level.

Lineage tracing. Embryos were injected with 100 pg of *Kaede* mRNA and 100 pg of *mGFP* mRNA (for contrast) at the one-cell stage, developed in the dark until

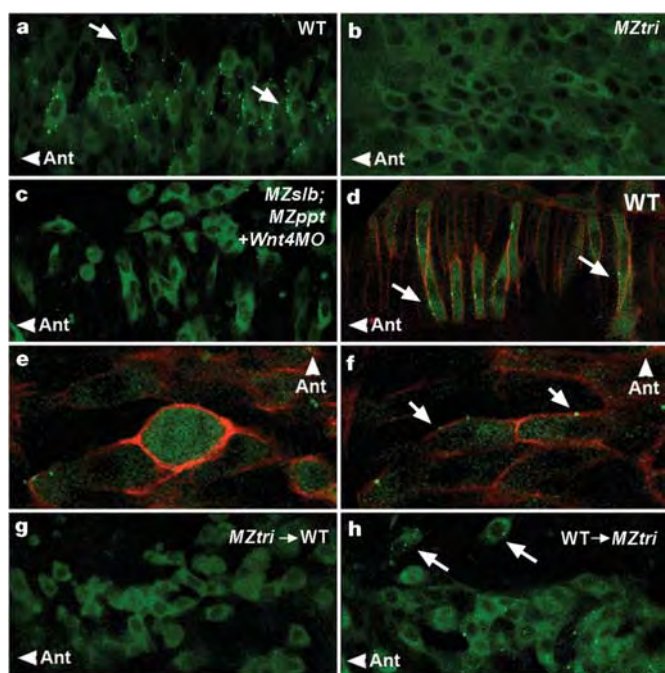


Figure 4 | Anterior membrane localization of Gfp-Pk as a marker of planar polarity. Confocal images taken at the level of the anterior spinal cord, through the dorsal–ventral plane of the neural keel (**a–c, e–h**) or notochord (**d**) of 8–10-somite-stage embryos. **a–c**, Scatter labelling of Gfp-Pk in the neural keel of a WT embryo (**a**), an *MZtri* mutant (**b**) or an *MZslb;MZppt* mutant (**c**) injected with 6 ng of *Wnt4* morpholino antisense oligonucleotides. **d**, Scatter labelling of Gfp-Pk plus mRFP in WT notochord, showing anterior membrane translocation of Gfp-Pk (arrows) in cells that undergo well-characterized convergence and extension movements in response to PCP signalling. **e, f**, A WT neural keel cell labelled with Gfp-Pk and mRFP, showing transient loss of polarity markers during mitosis (**e**) but re-establishment of membrane-localized Gfp-Pk in both daughter cells (**f**, arrows). **g, h**, Chimaeric analysis of the autonomy of PCP signalling, using Gfp-Pk as a marker of planar polarity. **g**, *MZtri* cells transplanted into WT host embryos do not localize Gfp-Pk to the membrane, because *Vangl2* is required for Pk translocation. **h**, WT cells transplanted into *MZtri* hosts show reduced membrane Gfp-Pk localization and abnormal polarity (arrows). Ant, anterior direction.

five-to-six-somite stages, and live-mounted in 0.8% agarose. The Kaede fluorophore was converted from green to red by focusing a 60-s pulse of ultraviolet light specifically on one lateral half of the neural plate, using the pinhole of a Zeiss LSM510 confocal microscope. Embryos were imaged both immediately and 10 h after treatment with ultraviolet light.

Received 6 September; accepted 26 October 2005.

- Copp, A. J., Greene, N. D. & Murdoch, J. N. The genetic basis of mammalian neurulation. *Nature Rev. Genet.* **4**, 784–793 (2003).
- Greene, N. D., Gerrelli, D., Van Straaten, H. W. & Copp, A. J. Abnormalities of floor plate, notochord and somite differentiation in the loop-tail (Lp) mouse: a model of severe neural tube defects. *Mech. Dev.* **73**, 59–72 (1998).
- Kibar, Z. *et al.* Ltap, a mammalian homolog of *Drosophila* Strabismus/Van Gogh, is altered in the mouse neural tube mutant Loop-tail. *Nature Genet.* **28**, 251–255 (2001).
- Murdoch, J. N., Doudney, K., Paternotte, C., Copp, A. J. & Stanier, P. Severe neural tube defects in the loop-tail mouse result from mutation of *Lp1*, a novel gene involved in floor plate specification. *Hum. Mol. Genet.* **10**, 2593–2601 (2001).
- Keller, R. Shaping the vertebrate body plan by polarized embryonic cell movements. *Science* **298**, 1950–1954 (2002).
- Wallingford, J. B., Fraser, S. E. & Harland, R. M. Convergent extension: the molecular control of polarized cell movement during embryonic development. *Dev. Cell* **2**, 695–706 (2002).
- Goto, T. & Keller, R. The planar cell polarity gene strabismus regulates

- convergence and extension and neural fold closure in *Xenopus*. *Dev. Biol.* **247**, 165–181 (2002).
- Wallingford, J. B. & Harland, R. M. Neural tube closure requires Dishevelled-dependent convergent extension of the midline. *Development* **129**, 5815–5825 (2002).
 - Kimmel, C. B., Warga, R. M. & Kane, D. A. Cell cycles and clonal strings during formation of the zebrafish central nervous system. *Development* **120**, 265–276 (1994).
 - Concha, M. L. & Adams, R. J. Oriented cell divisions and cellular morphogenesis in the zebrafish gastrula and neurula: a time-lapse analysis. *Development* **125**, 983–994 (1998).
 - Geldmacher-Voss, B., Reugels, A. M., Pauls, S. & Campos-Ortega, J. A. A 90-degree rotation of the mitotic spindle changes the orientation of mitoses of zebrafish neuroepithelial cells. *Development* **130**, 3767–3780 (2003).
 - Park, M. & Moon, R. T. The planar cell-polarity gene *stbm* regulates cell behaviour and cell fate in vertebrate embryos. *Nature Cell Biol.* **4**, 20–25 (2002).
 - Jessen, J. R. *et al.* Zebrafish trilobite identifies new roles for Strabismus in gastrulation and neuronal movements. *Nature Cell Biol.* **4**, 610–615 (2002).
 - Ciruna, B. *et al.* Production of maternal-zygotic mutant zebrafish by germ-line replacement. *Proc. Natl Acad. Sci. USA* **99**, 14919–14924 (2002).
 - Heisenberg, C. P. *et al.* Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* **405**, 76–81 (2000).
 - Rauch, G. J. *et al.* Wnt5 is required for tail formation in the zebrafish embryo. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 227–234 (1997).
 - Matsui, T. *et al.* Noncanonical Wnt signaling regulates midline convergence of organ primordia during zebrafish development. *Genes Dev.* **19**, 164–175 (2005).
 - Wallingford, J. B. & Harland, R. M. *Xenopus* Dishevelled signaling regulates both neural and mesodermal convergent extension: parallel forces elongating the body axis. *Development* **128**, 2581–2592 (2001).
 - Thisse, C. & Thisse, B. Antivin, a novel and divergent member of the TGF β superfamily, negatively regulates mesoderm induction. *Development* **126**, 229–240 (1999).
 - Meno, C. *et al.* Mouse Lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. *Mol. Cell* **4**, 287–298 (1999).
 - Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H. & Miyawaki, A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc. Natl Acad. Sci. USA* **99**, 12651–12656 (2002).
 - Strutt, D. I. The asymmetric subcellular localisation of components of the planar polarity pathway. *Semin. Cell Dev. Biol.* **13**, 225–231 (2002).
 - Jiang, D., Munro, E. M. & Smith, W. C. Ascidian prickle regulates both mediolateral and anterior–posterior cell polarity of notochord cells. *Curr. Biol.* **15**, 79–85 (2005).
 - Jenny, A., Darken, R. S., Wilson, P. A. & Mlodzik, M. Prickle and Strabismus form a functional complex to generate a correct axis during planar cell polarity signaling. *EMBO J.* **22**, 4409–4420 (2003).
 - Lyons, D. A. *et al.* *erbb3* and *erbb2* are essential for Schwann cell migration and myelination in zebrafish. *Curr. Biol.* **15**, 513–524 (2005).
 - Gong, Y., Mo, C. & Fraser, S. E. Planar cell polarity signalling controls cell division orientation during zebrafish gastrulation. *Nature* **430**, 689–693 (2004).
 - Djiane, A., Yogev, S. & Mlodzik, M. The apical determinants aPKC and dPati regulate Frizzled-dependent planar cell polarity in the *Drosophila* eye. *Cell* **121**, 621–631 (2005).
 - van Straaten, H. W. & Copp, A. J. Curly tail: a 50-year history of the mouse spina bifida model. *Anat. Embryol. (Berl.)* **203**, 225–237 (2001).
 - Gritsman, K. *et al.* The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. *Cell* **97**, 121–132 (1999).
 - Carreira-Barbosa, F. *et al.* Prickle 1 regulates cell movements during gastrulation and neuronal migration in zebrafish. *Development* **130**, 4037–4046 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Talbot and D. Lyons for sharing their protocol for pharmacological inhibition of cell division, A. Chitnis for useful discussion, and L. Solnica-Krezel, W. Talbot, J. Wallingford, A. Giraldez, D. Prober and J. Rihel for comments on the manuscript. This work was supported by grants from the NIH to A.F.S. and M.M. A.F.S. was a Scholar of the McKnight Foundation for Neuroscience, an Irma T. Hirsch Trust Career Scientist and an Established Investigator of the American Heart Association. B.C. was supported by a long-term fellowship from the Human Frontier Science Program.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.C. (ciruna@sickkids.ca) or A.F.S. (schier@fas.harvard.edu).

A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates

Daniel F. Jarosz^{1*}, Veronica G. Godoy^{2*}, James C. Delaney^{1,3}, John M. Essigmann^{1,3} & Graham C. Walker²

Translesion synthesis (TLS) by Y-family DNA polymerases is a chief mechanism of DNA damage tolerance¹. Such TLS can be accurate or error-prone, as it is for bypass of a cyclobutane pyrimidine dimer by DNA polymerase η (XP-V or Rad30) or bypass of a (6-4) TT photoproduct by DNA polymerase V (UmuD₂C), respectively^{2,3}. Although DinB is the only Y-family DNA polymerase conserved among all domains of life, the biological rationale for this striking conservation has remained enigmatic⁴. Here we report that the *Escherichia coli* *dinB* gene is required for resistance to some DNA-damaging agents that form adducts at the N²-position of deoxyguanosine (dG). We show that DinB (DNA polymerase IV) catalyses accurate TLS over one such N²-dG adduct (N²-furfuryl-dG), and that DinB and its mammalian orthologue, DNA polymerase κ , insert deoxycytidine (dC) opposite N²-furfuryl-dG with 10–15-fold greater catalytic proficiency than opposite undamaged dG. We also show that mutating a single amino acid, the ‘steric gate’ residue of DinB (Phe13 → Val) and that of its archaeal homologue Dbh (Phe12 → Ala), separates the abilities of these enzymes to perform TLS over N²-dG adducts from their abilities to replicate an undamaged template. We propose that DinB and its orthologues are specialized to catalyse relatively accurate TLS over some N²-dG adducts that are ubiquitous in nature, that lesion bypass occurs more efficiently than synthesis on undamaged DNA, and that this specificity may be achieved at least in part through a lesion-induced conformational change.

Although DinB is strongly upregulated as part of the SOS DNA damage response and *dinB*⁺ function has been implicated in untargeted mutagenesis of λ phage, adaptive mutagenesis and –1 frameshift mutagenesis when *dinB*⁺ is overexpressed in exponential phase^{5–9}, these phenotypes seem inadequate to account for the strong conservation of the DinB subfamily of DNA polymerases during evolution. We therefore exposed an *E. coli* strain bearing a precise deletion of the *dinB* gene to various DNA-damaging agents to gain insights into DinB function *in vivo*. The Δ *dinB* strain shows a marked sensitivity to nitrofurazone (NFZ; Supplementary Fig. S1a) that can be complemented *in trans* by *dinB*⁺ under its native promoter on a low copy-number plasmid (see Fig. 3a). The killing curve of a Δ umuC strain is indistinguishable from that of wild type (Fig. S1a), indicating that DinB is responsible for most TLS over potentially lethal NFZ-induced adducts. The Δ *dinB* mutant also shows increased sensitivity to killing by 4-nitroquinoline-1-oxide (4-NQO; Supplementary Fig. S1b and see Fig. 3b), but in this case TLS by UmuD₂C makes a contribution to survival in a *dinB*⁺ background (Supplementary Fig. S1b). Deletion of *polB*, which encodes DNA polymerase II (pol II) and is also induced by the SOS response¹⁰, does not increase sensitivity to either agent (data not shown).

Before forming stable N²-dG adducts *in vivo*, nitrofurans such as

NFZ must be reduced and acetylated¹¹. Likewise, at least half of the adducts that 4-NQO produces are N²-dG adducts^{11,12}. To address whether the NFZ resistance of a *dinB*⁺ strain arises from N²-dG lesion bypass, wild-type DinB was expressed and purified from *E. coli* (Supplementary Fig. S2), and oligonucleotide substrates were constructed that contained a site-specific N²-furfuryl-dG (Supplementary Fig. S3a), a structural analogue of the principal N²-dG adduct formed by NFZ. Whereas *E. coli* DNA pol I is strongly blocked by this lesion (Fig. 1a), DinB has markedly different properties. In the presence of all four deoxyribonucleotide triphosphates, DinB shows an increased catalytic proficiency on the N²-furfuryl-dG template relative to an undamaged template (Fig. 1b). Standing-start¹³ experiments (see Supplementary Fig. S4) indicate that DinB is 15-fold more proficient at adding dC opposite N²-furfuryl-dG than opposite undamaged dG (Fig. 1c). DNA pol κ , the mammalian DinB orthologue, is also considerably more proficient at adding dC opposite N²-furfuryl-dG than undamaged dG (Fig. 1d), indicating that this striking specificity has been conserved in eukaryotes. Furthermore, DinB bypass of N²-furfuryl-dG is not only proficient, but also accurate (Fig. 1e). This is achieved in part from a preference for correct dC insertion and in part from a preference for elongating from dC correctly paired with N²-furfuryl-dG (Table 1).

These observations suggest that a physiological role of DinB and its orthologues is to catalyse accurate TLS over some N²-dG adducts. This hypothesis received strong support from our construction of a *dinB* mutant that almost eliminates the ability of DinB to perform this type of TLS without impairing its ability to replicate undamaged DNA. We designed the *dinB* mutant after constructing a homology model of DinB encountering an N²-furfuryl-dG lesion based on the structure of *Sulfolobus solfataricus* Dpo4^{14,15} (Supplementary Fig. S5a–d). We noted a pocket in the enzyme next to the template base that could potentially accommodate the N²-furfuryl-dG adduct, bringing it into proximity with Phe 13. This residue corresponds to Phe 12 of the *S. acidocaldarius* DinB homologue (Dbh), the ‘steric gate’ that prevents the improper incorporation of ribonucleotide substrates by that enzyme¹⁶. Speculating that an active site rearrangement involving the N²-furfuryl-dG adduct, the Phe 13 steric gate residue and the incoming nucleotide might favour catalysis, we mutated the planar hydrophobic Phe 13 steric gate to a sterically different but still hydrophobic valine residue.

Primer extension assays using DinB(F13V), which purified to homogeneity indistinguishably from wild-type DinB (Supplementary Fig. S6), showed that DinB(F13V) is almost unable to carry out TLS over N²-furfuryl-dG, although its activity on undamaged DNA is largely unaffected (Table 1 and Fig. 2a). The F13V mutation has a modest effect on the ability of DinB to discriminate against ribonucleotides, increasing the frequency of their misincorporation from <10^{–5} (limit of detection) to ~10^{–3}. Because the steric gates of all

¹Department of Chemistry, ²Department of Biology, and ³Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

*These authors contributed equally to this work.

DinB orthologues are phenylalanine or tyrosine residues, we considered whether the corresponding mutation in these enzymes would likewise separate their TLS activities from their ability to replicate undamaged templates. We therefore assayed the archaeal DinB orthologue Dbh and its steric gate mutant Dbh(F12A)¹⁶ on N^2 -furfuryl-dG and undamaged templates. Whereas wild-type Dbh replicates both templates with comparable efficiencies at 37 °C, the F12A derivative shows disproportionately reduced activity on the damaged template (Fig. 2b, c).

To determine whether the F13V mutation specifically eliminates N^2 -dG lesion bypass without affecting other properties of DinB, we examined bypass of two other well-studied lesions, (+)-*trans*-anti-benzo[a]pyrene- N^2 -dG (N^2 -B[a]P-dG; Supplementary Fig. S3b) and a tetrahydrofuran abasic site analogue^{17,18} (Supplementary Fig. S3c). Although DinB-catalysed bypass of the N^2 -B[a]P-dG lesion is inefficient¹⁷ as compared with bypass of N^2 -furfuryl-dG, the F13V mutation similarly eliminates its ability to perform this type of TLS (Fig. 2d). Furthermore, like the wild-type enzyme¹⁸, DinB(F13V) is unable to bypass a tetrahydrofuran abasic site analogue efficiently (Supplementary Fig. S7), indicating that the F13V mutation has not relaxed the specificity of DinB *in vitro*. Although it is possible that the F13V mutation also affects DinB bypass of some other lesion, these data indicate that it specifically eliminates bypass of N^2 -dG lesions.

To establish whether N^2 -dG lesion bypass is required for *dinB*-dependent resistance to NFZ and 4-NQO, we examined the ability of a low-copy number plasmid carrying the *dinB*(F13V) allele under its own promoter to complement a Δ *dinB* strain for NFZ and 4-NQO resistance (Fig. 3a, b). Although the mutant protein is expressed from this plasmid *in vivo* (data not shown), *pdinB*(F13V) is unable to complement NFZ or 4-NQO resistance, which is consistent with the notion that an N^2 -dG adduct is responsible for NFZ lethality. Furthermore, *pdinB*(F13V) exacerbates the sensitivity of the Δ *dinB*

strain to these agents, to an even greater degree than a plasmid encoding a catalytically inactive DinB(D103N) mutant protein (*pdinB003*)¹⁹ (Fig. 3a, b), suggesting that it is interfering with some cellular process that can otherwise contribute modestly to NFZ resistance. The plasmid-borne *dinB*(F13V) allele does not affect viability of the *dinB*⁺ strain, but it has a dominant negative effect on survival after treatment with either NFZ or 4-NQO (Supplementary Fig. S8). We conclude that this dominance is largely due to an impairment of TLS rather than ribonucleotide misincorporation into DNA by DinB(F13V) because dominance is still observed in an *rnhB* mutant²⁰ (Supplementary Fig. S8) and the mutant enzyme still favours dNTP incorporation *in vitro* (Table 1). Taken together, our data indicate that the aromatic steric gate residue of DinB is required for TLS over N^2 -dG adducts both *in vivo* and *in vitro*.

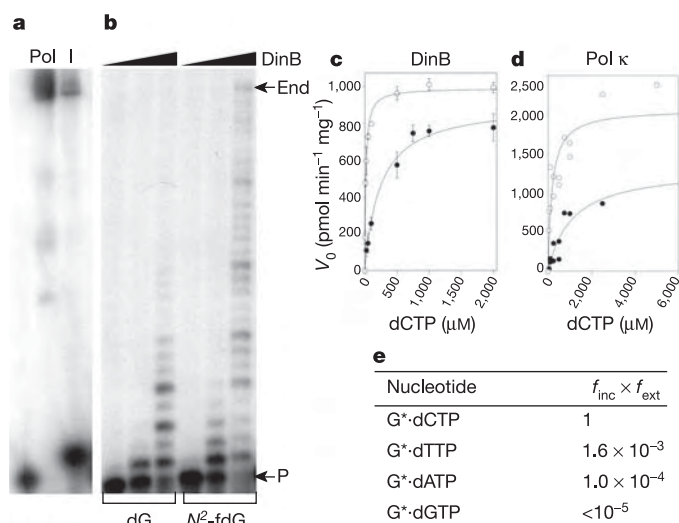


Figure 1 | Bypass of N^2 -furfuryl-dG. **a**, Primer (lane 1) extension products of *E. coli* pol I (5 nM) on undamaged dG (lane 2, 95.3% extension) and N^2 -furfuryl-dG-damaged templates (lane 3, 8.2% extension; see Supplementary Information). **b**, Running-start primer (P) extension reactions with 1, 10 or 50 nM DinB protein and 250 μ M dNTPs. Lanes 1–3, undamaged dG template (0.03, 2.2 and 81.8% extension, respectively); lanes 4–6, N^2 -furfuryl-dG-damaged template (0.05, 65.5 and 91.1% extension, respectively). **c**, Plot of initial reaction velocity versus initial concentration of dCTP in standing-start assays on undamaged dG (filled circles) and N^2 -furfuryl-dG-damaged templates (open circles). Error bars represent 1 s.d. determined from three reactions. **d**, As **c**, but with the mammalian DinB orthologue pol κ . **e**, Fidelity of DinB bypass of N^2 -furfuryl-dG measured by using standing-start incorporation and extension assays¹³. Error of these measurements is $\sim 20\%$.

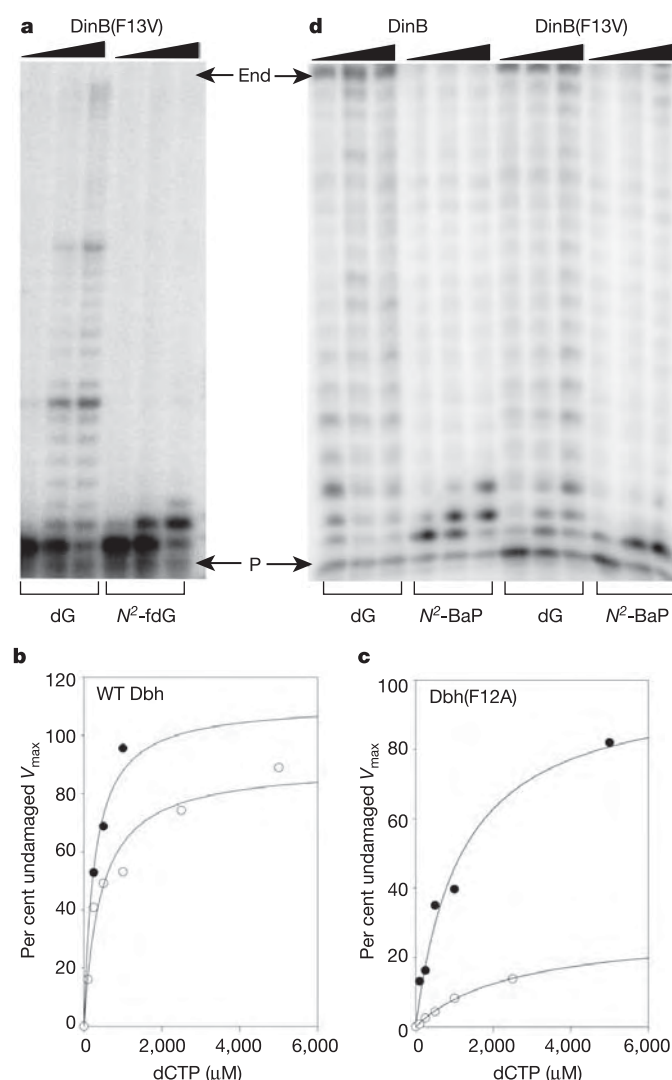


Figure 2 | A single mutation in DinB or its archaeal orthologue Dbh separates their TLS and DNA polymerase activities. **a**, Running start primer extension reactions using 1, 10 and 50 nM DinB(F13V) on undamaged dG (lanes 1–3, 0, 79.1 and 86.8% extension, respectively) and N^2 -furfuryl-dG-damaged templates (lanes 4–6, 0, 7.5 and 16.3% extension, respectively). DinB(F13V) retains DNA polymerase activity but is compromised for TLS. **b**, Plot of relative initial velocity versus initial dCTP concentration for Dbh on undamaged dG (filled circles) and N^2 -furfuryl-dG (open circles). **c**, As **b** but for Dbh(F12A); activity is disproportionately reduced on the N^2 -furfuryl-dG-damaged template. **d**, Running-start primer (P) extension assays with 1, 10 and 50 nM DinB or DinB(F13V) on undamaged and N^2 -B[a]P-dG-damaged templates. Lanes 1–12 show 29.8, 62, 88.7, 1.3, 17.8, 32.4, 11.5, 35.5, 53.8, 0, 0 and 0.2% extension, respectively.

Table 1 | Kinetic parameters for N^2 -furfuryl-dG bypass, mismatch extension and rNTP insertion

Substrate†	Enzyme	$V_{\max}$ (pmol min ⁻¹ mg ⁻¹)	K_m (μM)	$V_{\max}/K_m$ (pmol min ⁻¹ mg ⁻¹ M ⁻¹)‡	f_{inc}
dNTP insertion					
G-dCTP	WT DinB	910.3	240.1	3.8×10^6	1.0
G*-dCTP	WT DinB	992.8	16.1	6.2×10^7	16.3
G*-dTTP	WT DinB	155.5	36.7	4.2×10^6	1.1
G*-dATP	WT DinB	81.6	20.2	4.0×10^6	1.1
G*-dGTP	WT DinB	Undetectable	-	-	-
G-dCTP	DinB(F13V)	1461.0	117.7	1.2×10^7	1.0
G*-dCTP	DinB(F13V)	300.8	227.1	1.3×10^6	0.1
Mismatch extension					
G*C	WT DinB	807.7	2.4	3.4×10^8	1.0
G*T	WT DinB	101.2	14.2	8.2×10^6	2.4×10^{-2}
G*A	WT DinB	46.4	86.2	5.4×10^5	1.6×10^{-3}
rCTP insertion					
G-rCTP	WT DinB	Undetectable	-	-	-
G*-rCTP	WT DinB	28.3	932.0	3.0×10^4	N/A
G-rCTP	DinB(F13V)	23.7	325.0	7.3×10^4	N/A
G*-rCTP	DinB(F13V)	8.1	293.0	2.8×10^4	N/A
Template parameters					
Undamaged dG	WT DinB	274.0	13.8 ± 1.7	2.0×10^7	1.0
N^2 -furfuryl-dG	WT DinB	1,040.9	20.5 ± 5.2	5.1×10^7	2.6

†G* represents N^2 -furfuryl-dG; the s.e.m. is ~20%.‡ $V_{\max}$ and K_m for the DNA substrates reveal preferential activity on N^2 -furfuryl-dG.

Does DinB prevent mutagenesis caused by NFZ or 4-NQO *in vivo*, or does it promote mutagenesis, a behaviour frequently attributed to Y-family DNA polymerases? Loss of $dinB^+$ does not alter the frequency of NFZ-induced rifampicin-resistant (Rif^r) mutations (mean $\pm$ 1 s.d., $29.6 \pm 9.7 \times 10^{-9}$ for $dinB^+$ versus $31.4 \pm 11.8 \times 10^{-9}$ for $\Delta dinB$). Given the markedly greater ability of a $dinB^+$ strain to survive NFZ treatment as compared with a $\Delta dinB$ strain, this indicates that DinB is not a mutagenic polymerase when bypassing NFZ-induced lesions that are lethal in its absence. Loss of $dinB^+$ function results in an increase in the frequency of Rif^r mutants

upon 4-NQO treatment ($48.1 \pm 1.7 \times 10^{-9}$ for $dinB^+$ versus $453 \pm 181 \times 10^{-9}$ for $\Delta dinB$). This implies that DinB carries out accurate TLS over a class of lethal 4-NQO-induced lesions that are bypassed in a more error-prone fashion in its absence, most likely by UmuD₂C (ref. 21). These data indicate that the $dinB$ gene product bypasses the lethal lesions generated by NFZ and 4-NQO with unexpectedly high fidelity *in vivo*, thus resembling the behaviour of DNA pol η when bypassing cyclobutane pyrimidine dimers²².

The *in vitro* data discussed above indicate that the F13V mutation almost eliminates the ability of DinB to bypass N^2 -furfuryl-dG and does not relax its specificity with respect to the lesions that it can bypass, but does DinB(F13V) replicate undamaged DNA with reduced fidelity? Using a set of $\Delta dinB$ strains carrying various plasmid-borne $dinB$ alleles (Fig. 3a, b), we examined the frequencies of spontaneous and NFZ-induced mutation to Rif^r. We observed no increase in the frequency of spontaneous or NFZ-induced Rif^r mutations between the $dinB^+$ and $dinB(F13V)$ alleles (Fig. 3c), indicating that the F13V mutation does not result in DinB(F13V) becoming a mutator polymerase. We also compared the effect of $dinB(F13V)$ on spontaneous mutation to that of $dinB^+$ using derivatives of the strain CC102 (ref. 23). This strain carries a *lacZ* allele that reverts by a GC to AT transition, the most frequent DinB error that we detected *in vitro*. Here again, we detected no increase in Lac⁺ reversion between the $dinB(F13V)$ derivative ($11 \pm 5 \times 10^{-9}$) and that of the $dinB^+$ strain ($8 \pm 5 \times 10^{-9}$), indicating that the F13V mutation does not decrease the fidelity of DinB.

The highly conserved steric gate residue, Phe13, clearly has a crucial role in bypass of N^2 -dG adducts by DinB, but further work will be required to establish whether it participates in an N^2 -dG lesion-induced conformational change that permits preferential replication of this type of damaged DNA template. Nevertheless, some of our observations are consistent with such a lesion-induced conformational change (Table 1) including, first, the ability of wild-type DinB to incorporate detectably low levels of rNTPs only when acting on the N^2 -furfuryl-dG bearing template; second, no detectable increase in rNTP incorporation by DinB(F13V) on the N^2 -furfuryl-dG bearing template relative to an undamaged control; and last, a lower apparent Michaelis constant (K_m) for dCTP when DinB is bound to an N^2 -furfuryl-dG standing-start template rather than to the corresponding dG template, coupled with a higher apparent maximal enzyme-catalysed reaction velocity ($V_{\max}$) for the damaged primer or template itself.

DinB may have a role as a mutator polymerase under certain conditions of biological stress or in some sequence contexts^{7,19}. Because other amino acids can act as steric gates in other DNA

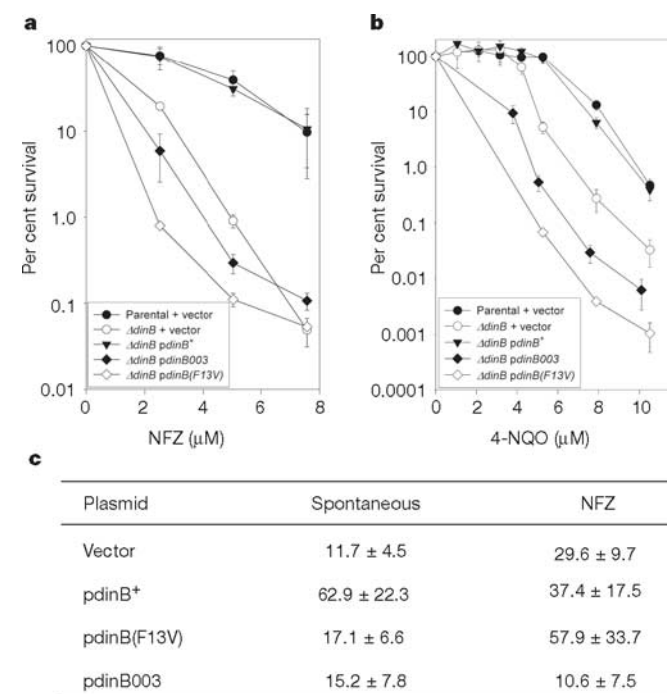


Figure 3 | Importance of DinB F13 residue *in vivo*. **a**, Unlike $pdinB^+$, $pdinB(F13V)$ is unable to restore NFZ resistance in the $\Delta dinB$ strain but instead exacerbates the sensitivity, like $pdinB003$, which encodes a catalytically inactive DinB(D103N) protein. Error bars indicate 1 s.d. determined from three experiments. **b**, $pdinB(F13V)$ and $pdinB003$ also exacerbate the sensitivity of the $\Delta dinB$ strain to 4-NQO. **c**, Spontaneous and induced mutation frequencies per 10^9 bacteria to Rif^r.

polymerases²⁴, however, the evolutionary conservation of this aromatic steric gate residue with a crucial role in TLS suggests that N^2 -dG adduct bypass is an important and physiologically relevant property of the DinB subfamily of Y-family DNA polymerases. Present at levels comparable to those of 8-oxo-G *in vivo*, N^2 -dG adducts are formed from byproducts of diverse cellular processes, including lipid peroxidation²⁵. Furthermore, there is evidence that these minor groove adducts may be recalcitrant to excision repair²⁶. In mammalian cells, the N^2 -dG adduct of the carcinogen acetylaminofluorene persists even though it is the least common dG isomer formed²⁷. Finally, our results emphasize that Y-family DNA polymerases, although perhaps relatively error-prone under some conditions^{7,8,10,19}, can also be specialized for proficient and accurate replication of a particular class of damaged DNA.

METHODS

Sensitivity and mutation frequency determination. Details of strain and plasmid construction are included as Supplementary Information. From single colonies, strains were grown for 16 h to saturation, diluted 1:1,000 into LB medium and then grown to 5×10^9 c.f.u. per ml. From this freshly saturated culture, dilutions were plated on LB agar containing ampicillin and 0–15 μ M NFZ or 4-NQO. Stock solutions of NFZ and 4-NQO were freshly prepared in *N,N*-dimethyl formamide. About ten of these colonies were suspended in M9 salts and deposited on plates containing rifampicin (100 μ g ml⁻¹) to determine the number of Rif^r mutants. This number was corrected for the number of viable cells in each colony. We measured GC to AT transitions using a Δ *dinB* derivative of the CC102 strain²³.

Template synthesis and construction. The N^2 -furfuryl-dG adduct was made by a postsynthetic derivatization approach²⁸, described in detail in the Supplementary Information. MALDI-TOF mass spectrometry of the purified oligonucleotide gave a mass of 4,986.26 (4,986.27 calculated) for the single, negatively charged moiety (Supplementary Fig. S9). The 16-nucleotide lesion-bearing oligomer was ligated to 5'-GGTTACTCAGATCAGGCCTGC-3' at the 5' end and 5'-GGCTGCAGCTGTACTATCATATGC-3' at the 3' end by standard protocols, and gel-purified to remove the ligation scaffolds 5'-AGGTCCTCG-CAGGCCTGA-3' and 5'-CAGTCGCAGCCGACGCC-3'. The benzo[a]pyrene lesion is in the sequence context 5'-GACTACGTACTGTACATXCACAGCCTATCTGGCCAGATCCGC-3'.

Primer extension assays. Details of the protein purification procedure are given in the Supplementary Information. Assays were performed and quantified by using either standing or running start primers of the sequences 5'-GCATATGATAGTACAGCTGCAGCCGGACGCC-3' or 5'-GCATATGATAGTACAGCTGCAGCCGGACGCC-3', respectively, for all templates except the N^2 -B[a]P-dG-bearing substrate¹³. For that substrate, the primer 5'-GCGGATCTGGCCAGATAGCGTGT-3' (running) was used. In brief, assays were conducted in a 10- μ l volume containing 1, 10 or 50 nM DinB, 10 nM primer and template, either 250 μ M dNTPs or 0–2,000 μ M dCTP, 50 mM HEPES (pH 7.5), 100 mM KCl, 7.5 mM MgCl₂, 5% glycerol and 0.1% bovine serum albumin. Reactions were initiated with dNTPs and quenched after incubation for 15 min (or as noted in the figure legends) at 37 °C. The per cent extension is defined as the percentage of primers that are extended past the lesion. The same conditions were used to assay Dbh and pol κ , except that the enzyme concentration was 10 nM and 2 nM, respectively. Products were analysed on a 12% denaturing polyacrylamide gel and quantified on a phosphorimager (Molecular Dynamics).

Received 1 July; accepted 11 October 2005.

1. Friedberg, E. C., Wagner, R. & Radman, M. Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science* **296**, 1627–1630 (2002).
2. Washington, M. T., Johnson, R. E., Prakash, L. & Prakash, S. Accuracy of lesion bypass by yeast and human DNA polymerase η . *Proc. Natl Acad. Sci. USA* **98**, 8355–8360 (2001).
3. Tang, M. *et al.* Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature* **404**, 1014–1018 (2000).
4. Fuchs, R. P., Fujii, S. & Wagner, J. Properties and functions of *Escherichia coli*: pol IV and pol V. *Adv. Protein Chem.* **69**, 229–264 (2004).
5. Kenyon, C. J. & Walker, G. C. DNA-damaging agents stimulate gene expression at specific loci in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **77**, 2819–2823 (1980).
6. Brotcorne-Lannoye, A. & Maenhaut-Michel, G. Role of RecA protein in untargeted UV mutagenesis of bacteriophage λ : evidence for the requirement for the *dinB* gene. *Proc. Natl Acad. Sci. USA* **83**, 3904–3908 (1986).
7. McKenzie, G. J., Lee, P. L., Lombardo, M. J., Hastings, P. J. & Rosenberg, S. M. SOS mutator DNA polymerase IV functions in adaptive mutation and not adaptive amplification. *Mol. Cell* **7**, 571–579 (2001).

8. Kim, S. R. *et al.* Multiple pathways for SOS-induced mutagenesis in *Escherichia coli*: an overexpression of *dinB/dinP* results in strongly enhancing mutagenesis in the absence of any exogenous treatment to damage DNA. *Proc. Natl Acad. Sci. USA* **94**, 13792–13797 (1997).
9. Kim, S. R., Matsui, K., Yamada, M., Gruz, P. & Nohmi, T. Roles of chromosomal and episomal *dinB* genes encoding DNA pol IV in targeted and untargeted mutagenesis in *Escherichia coli*. *Mol. Genet. Genomics* **266**, 207–215 (2001).
10. Goodman, M. F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* **71**, 17–50 (2002).
11. Whiteway, J. *et al.* Oxygen-insensitive nitroreductases: analysis of the roles of *nfsA* and *nfsB* in development of resistance to 5-nitrofur derivatives in *Escherichia coli*. *J. Bacteriol.* **180**, 5529–5539 (1998).
12. Panigrahi, G. B. & Walker, I. G. The N^2 -guanine adduct but not the C8-guanine or N^6 -adenine adducts formed by 4-nitroquinoline 1-oxide blocks the 3'–5' exonuclease action of T4 DNA polymerase. *Biochemistry* **29**, 2122–2126 (1990).
13. Creighton, S., Bloom, L. B. & Goodman, M. F. Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Methods Enzymol.* **262**, 232–256 (1995).
14. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).
15. Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **107**, 91–102 (2001).
16. DeLucia, A. M., Grindley, N. D. & Joyce, C. M. An error-prone family Y DNA polymerase (DinB homolog from *Sulfolobus solfataricus*) uses a 'steric gate' residue for discrimination against ribonucleotides. *Nucleic Acids Res.* **31**, 4129–4137 (2003).
17. Shen, X. *et al.* Efficiency and accuracy of SOS-induced DNA polymerases replicating benzo[a]pyrene-7,8-diol 9,10-epoxide A and G adducts. *J. Biol. Chem.* **277**, 5265–5274 (2002).
18. Gruz, P. *et al.* Synthetic activity of Sso DNA polymerase Y1, an archaeal DinB-like DNA polymerase, is stimulated by processivity factors proliferating cell nuclear antigen and replication factor C. *J. Biol. Chem.* **276**, 47394–47401 (2001).
19. Wagner, J. *et al.* The *dinB* gene encodes a novel *E. coli* DNA polymerase, DNA pol IV, involved in mutagenesis. *Mol. Cell* **4**, 281–286 (1999).
20. Rydberg, B. & Game, J. Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proc. Natl Acad. Sci. USA* **99**, 16654–16659 (2002).
21. Watanabe, M., Nohmi, T. & Ohta, T. Effects of the *umuDC*, *mucAB*, and *samAB* operons on the mutational specificity of chemical mutagenesis in *Escherichia coli*: II. Base substitution mutagenesis. *Mutat. Res.* **314**, 39–49 (1994).
22. Maher, V. M., Ouellette, L. M., Curren, R. D. & McCormick, J. J. Frequency of ultraviolet light-induced mutations is higher in xeroderma pigmentosum variant cells than in normal human cells. *Nature* **261**, 593–595 (1976).
23. Cupples, C. G. & Miller, J. H. A set of *lacZ* mutations in *Escherichia coli* that allow rapid detection of each of the six base substitutions. *Proc. Natl Acad. Sci. USA* **86**, 5345–5349 (1989).
24. Astatke, M., Ng, K., Grindley, N. D. & Joyce, C. M. A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl Acad. Sci. USA* **95**, 3402–3407 (1998).
25. Marnett, L. J. Oxy radicals, lipid peroxidation and DNA damage. *Toxicology* **181–182**, 219–222 (2002).
26. Zewail-foote, M. *et al.* The inefficiency of incisions of ecteinascidin 743–DNA adducts by the UvrABC nuclease and the unique structural feature of the DNA adducts can be used to explain the repair-dependent toxicities of this antitumor agent. *Chem. Biol.* **8**, 1033–1049 (2001).
27. Yasui, M. *et al.* Mutagenic properties of 3-(deoxyguanosin- N^2 -yl)-2-acetylaminofluorene, a persistent acetylaminofluorene-derived DNA adduct in mammalian cells. *Biochemistry* **43**, 15005–15013 (2004).
28. DeCorte, B. L. *et al.* Improved strategies for postoligomerization synthesis of oligodeoxynucleotides bearing structurally defined adducts at the N^2 position of deoxyguanosine. *Chem. Res. Toxicol.* **9**, 630–637 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Joyce and A. DeLucia for Dbh and Dbh(F12A); E. Friedberg, J. Stirman and P. Fischhaber for *Mus musculus* pol κ ; T. Nohmi for pYG768 and pYG782; N. Geacintov for benzo[a]pyrene-damaged substrate; A. Banerjee for assistance with Supplementary Fig. S5; and J. Tuttle and J. Bowers for advice on protein purification. This work was supported by the NIH (grants to G.C.W. and J.M.E.).

Author Contributions D.F.J. performed the protein purification and lesion bypass assays, and proposed the involvement of the steric gate residue in TLS. V.G.G. discovered the sensitivity of a Δ *dinB* strain to NFZ and 4-NQO and performed the mutagenesis experiments. J.C.D. constructed and purified the N^2 -furfuryl-dG-containing oligonucleotide substrate.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.C.W. (gwalker@mit.edu).

Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk

Peijian Zou^{1*}, Nikos Pinotsis^{1,2*}, Stephan Lange^{3,4}, Young-Hwa Song¹, Alexander Popov¹, Irene Mavridis², Olga M. Mayans^{1,5}, Mathias Gautel³ & Matthias Wilmanns¹

The Z-disk of striated and cardiac muscle sarcomeres is one of the most densely packed cellular structures in eukaryotic cells¹. It provides the architectural framework for assembling and anchoring the largest known muscle filament systems by an extensive network of protein–protein interactions, requiring an extraordinary level of mechanical stability. Here we show, using X-ray crystallography, how the amino terminus of the longest filament component, the giant muscle protein titin, is assembled into an antiparallel (2:1) sandwich complex by the Z-disk ligand telethonin. The pseudosymmetric structure of telethonin mediates a unique palindromic arrangement of two titin filaments, a type of molecular assembly previously found only in protein–DNA complexes. We have confirmed its unique architecture *in vivo* by protein complementation assays, and *in vitro* by experiments using fluorescence resonance energy transfer. The model proposed may provide a molecular paradigm of how major sarcomeric filaments are crosslinked, anchored and aligned within complex cytoskeletal networks.

The Z-disk of the sarcomere defines the lateral boundary of sarcomeric units within the myocyte cytoskeleton. In higher vertebrates it anchors and aligns at least three major sarcomeric filament systems, including actin, titin and nebulin^{1,2}. It also harbours many smaller protein components, some of which, including α -actinin and telethonin, have been mapped to distinct binding sites at the N terminus of titin^{3–7}. Their presence, proper sorting and localization within the Z-disk region are critical for myofibril assembly and for the maintenance of an intact Z-disk structure⁸. Components of the Z-disk are also involved in signalling processes that may regulate muscle development and degradation, as well as in linking contractile functions of muscle sarcomeres to membrane systems such as the sarcoplasmic reticulum or the T-tubules associated with excitation–contraction coupling^{2,8}.

The very N-terminal region of titin comprises a domain topology that has been predicted to consist of two immunoglobulin-like domains, referred to as Z1 and Z2 (ref. 9). Co-localization studies, two-hybrid interaction screens and pull-down assays have demonstrated that they interact with the N-terminal region of telethonin at the Z-disk periphery^{3,4}. However, although the interaction with telethonin has been considered as a ‘cap’ (hence the alternative name ‘T-cap’), or a ‘bolt’³, it is not known how titin–telethonin binding affects the overall architecture of myofibrils and their associated functions. The physiological importance of this interaction has been supported by evidence linking mutations in the N-terminal regions of titin, telethonin and the telethonin-binding site of muscle LIM protein (MLP) to different familial forms of

limb-girdle muscular dystrophy, as well as hypertrophic and dilated cardiomyopathy^{10–12}. These findings, together with complementary data from animal models, indicate the possible existence of a link between the titin–telethonin–MLP interaction and mechanical stress sensor pathways¹⁰.

Depending on the specific isoform of titin, 200–700 N-terminal residues of multimeric titin filaments are located within and cross over most of the Z-disk of striated muscle sarcomeres^{2,3,6}. However, no data are yet available on the molecular nature of the titin–titin association. Here we have determined the crystal structure of its N-terminal region in complex with the titin-binding domain of telethonin (Fig. 1, Table 1 and Supplementary Fig. S1). The latter domain is sufficient to localize telethonin to the Z-disk of cardiac myofibrils (Supplementary Fig. S2). Contrary to previous expectations³ and previous structural findings on other immunoglobulin-like domain-containing proteins (see Supplementary Information), our data reveal an antiparallel assembly of two titin molecules mediated by telethonin, indicating that telethonin might have a key role in titin assembly and Z-disk anchoring.

In the complex, the two N-terminal immunoglobulin-like domains of titin, Z1 and Z2 are in an extended conformation and are connected by a short three-residue linker. The two domains have similar structures (root-mean-square deviation (r.m.s.d.) = 0.66 Å, for all common main-chain atoms) and sequences (40 of 98 residues are identical). In each of the two titin molecules, domains Z1 and Z2 are almost equally translated by 48 Å and rotated by 53° and 61° with respect to each other, generating a superhelical coil arrangement of each titin N terminus. The second component of the complex, telethonin, forms a unique elongated structure with a central five-stranded antiparallel β -sheet that is extended by two exposed

Table 1 | Refinement statistics

Resolution (Å)	15.0–2.45
$R_{\text{work}}/R_{\text{free}}$ (%)	23.2/26.5
Number of atoms	
Protein	3,726
Ion	25
Water	179
B factors (Å ²)	
Protein	49.7
Ion	51.2
Water	48.6
R.m.s.d.	
Bond lengths (Å)	0.011
Bond angles (°)	1.370

¹EMBL-Hamburg c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany. ²Institute of Physical Chemistry, National Center for Scientific Research ‘Demokritos’, Aghia Paraskevi 15310, Athens, Greece. ³The Randall Division of Cell and Molecular Biophysics and Cardiovascular Division, King’s College London, Guy’s Campus, London SE1 1UL, UK. ⁴Institute of Cell Biology, ETH Zurich Hoenggerberg, CH-8093 Zurich, Switzerland. ⁵Biozentrum, University of Basel, Division of Structural Biology, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.

*These authors contributed equally to this work.

wing-shaped β -hairpin motifs (A–B, C–D). The two motifs are related by an approximate two-fold symmetry (rotation of 179.4°), generating a nearly perfect palindromic arrangement (Figs 1a and 2a–d). They can be superimposed with an r.m.s.d. of 0.38 Å, and in the corresponding structural alignment 11 of 19 matching residues are similar or identical.

The peculiar symmetry of the telethonin structure allows it to mediate titin–titin assembly (Fig. 1b). The four nearly identical titin–telethonin interfaces, I–IV, are formed from two intermolecular antiparallel β -sheets. Within each interface, the long, invariant β -strand G of one of two titin immunoglobulin-like domains interacts with one of the four β -strands (A, B, C or D) of the two wing-like hairpins from telethonin. Hence, each telethonin hairpin (A–B, C–D) provides the core for one of the two (3–2–3)-stranded, antiparallel titin–telethonin–titin β -sheets. The two β -sheets are separated by the telethonin core β -sheet that is flanked by the two short Z1–Z2 linkers of each sandwiching titin molecule. Taken together, the data on the titin–telethonin complex provide a novel

concept exemplifying how immunoglobulin-like domain-containing proteins may act as receptors for protein ligands, such as telethonin. In contrast to other immunoglobulin-like receptors^{13,14}, the protein–protein interface of the titin–telethonin sandwich complex is formed by main-chain-mediated intermolecular β -sheet interactions. In terms of general principles of complex formation of biological molecules, the titin–telethonin complex reveals an unprecedented analogy to palindromic or pseudopalindromic protein–DNA complexes¹⁵. Future data are needed to determine whether this type of palindromic complex is unique to the titin–telethonin interaction or whether it establishes a new principle of protein–protein interactions. The extensive network of interactions indicates that the N terminus of titin, after complex formation with telethonin, provides a rigid anchoring scaffold rather than adding to the molecular elasticity that has been observed in several titin segments². Indeed, such an anchoring structure that is resistant to external mechanical forces seems to be a prerequisite for elastic movements of other parts of titin, specifically within the sarcomeric I-band, under active muscle contraction–relaxation conditions, without the danger of uncontrolled disintegration. The bridging structure of telethonin indicates that it might be essential for the functional integrity of the titin filament in mature myofibrils.

To validate our structural data, we first designed an *in vitro* fluorescence resonance energy transfer (FRET) experiment, in which we introduced four site-specific donor–acceptor fluorophore pairs into the two titin molecules to measure residue–residue distances within the complex in solution (Fig. 3C). The FRET distances of all four donor–acceptor pairs well reflect those observed in the crystal structure (Fig. 3D). Because the FRET data can be neither modelled into a putative parallel titin–titin arrangement nor explained by other stoichiometries, they provide independent and unambiguous evidence for an antiparallel arrangement of the two titin molecules in the titin–telethonin complex in solution.

Subsequently, we performed two types of fluorescence imaging experiments to validate our structural findings of the antiparallel titin–telethonin complex under *in vivo* conditions. In the first approach, we used COS cells to test whether the titin–telethonin complex observed structurally can also form under *in vivo* conditions in the absence of a pre-existing sarcomeric filament system (Fig. 3A, B).

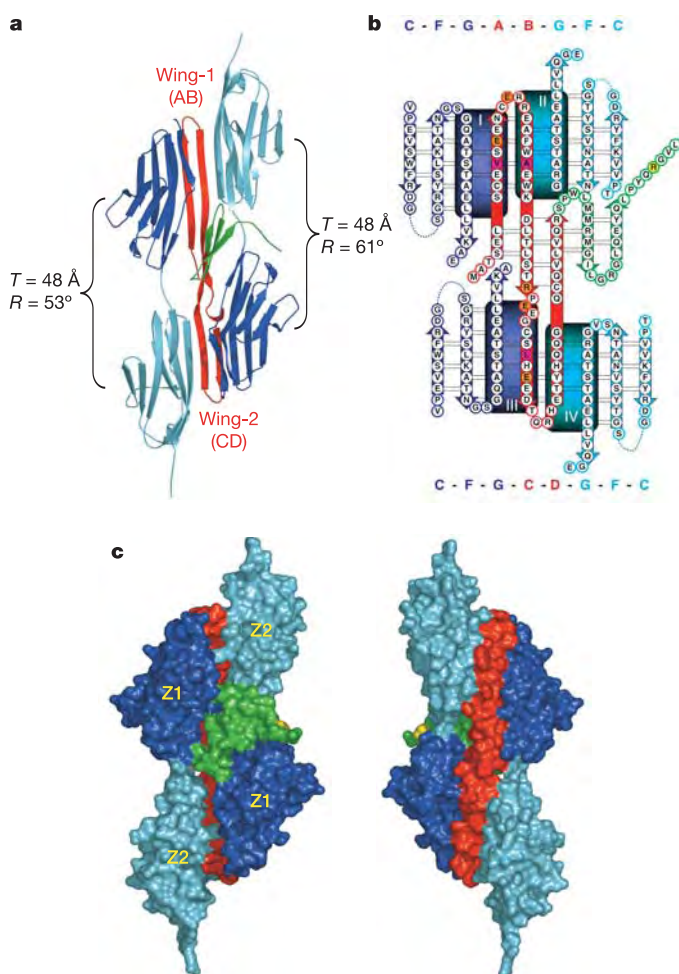


Figure 1 | Structure of the palindromic titin–telethonin–titin complex. Colour codes: blue, titin immunoglobulin-like domain Z1 (residues 1–98); cyan, titin immunoglobulin-like domain Z2, including the Z1–Z2 linker (99–196); red, telethonin(1–59); green, telethonin(60–90). **a**, Ribbon representation. **b**, Schematic representation of the β -sheet structure in the titin–telethonin (2:1) complex. Arg87 of telethonin, which is linked to dilated cardiomyopathy, is coloured yellow¹⁰. The β -sheet hydrogen bonds are depicted by lines. For clarity, only those parts of the two immunoglobulin-like tandem repeats from titin that are involved in interactions with telethonin are shown. **c**, Surface presentation of the titin–telethonin–titin complex in two orientations, rotated by 180° relative to each other.

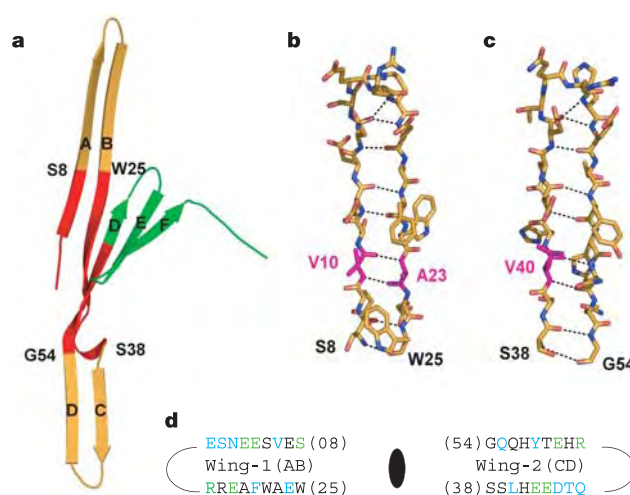


Figure 2 | Telethonin-mediated assembly and Z-disk anchoring of titin filaments. **a**, Ribbon presentation of the structure of telethonin, indicating the residue numbers of the β -hairpin wing boundaries. Colours as in Fig. 1. **b**, **c**, Stick presentations of the two β -hairpin wings of telethonin. Colours as in **a**, except those that have been mutated into prolines (pink) for validation purposes. **d**, Sequence representation of the two β -hairpin wings, indicating how the repeated sequence motif translates into a palindromic structural relation. Identical and similar residues are coloured in green and blue, respectively.

We employed a yellow fluorescence protein (YFP) reconstitution assay using two YFP half-domains (YN, YC)¹⁶. After transfection of COS cells with constructs resulting in a titin–telethonin complex, fluorescence from reconstituted YFP was detected only for the YN–titin–telethonin–titin–YC system in contrast to the YN–titin–telethonin–YC–titin system, thus fitting an antiparallel assembly of two titin molecules only (Fig. 3A). The data were confirmed by immunoblot assays of lysates from transfected COS cells (Fig. 3B).

To examine the correct targeting of telethonin to the endogenous titin N terminus within the sarcomeric Z-disk *in vivo*, we used

neonatal rat cardiac myocytes (NRCs), which express all sarcomeric components (Supplementary Fig. S2). To allow comparison of titin–telethonin binding under *in vitro* conditions, in a test cell line (COS cells) without sarcomeres and in sarcomere-containing muscle cells, we introduced several structure-based single-residue mutations in telethonin. In a first experiment series, we changed several residues involved in the titin–telethonin interface and tested their ability for titin–telethonin complex formation *in vitro* (Supplementary Fig. S3). None of them indicated abolition of the interaction, most probably because of the high stability of the complex formed. However, three

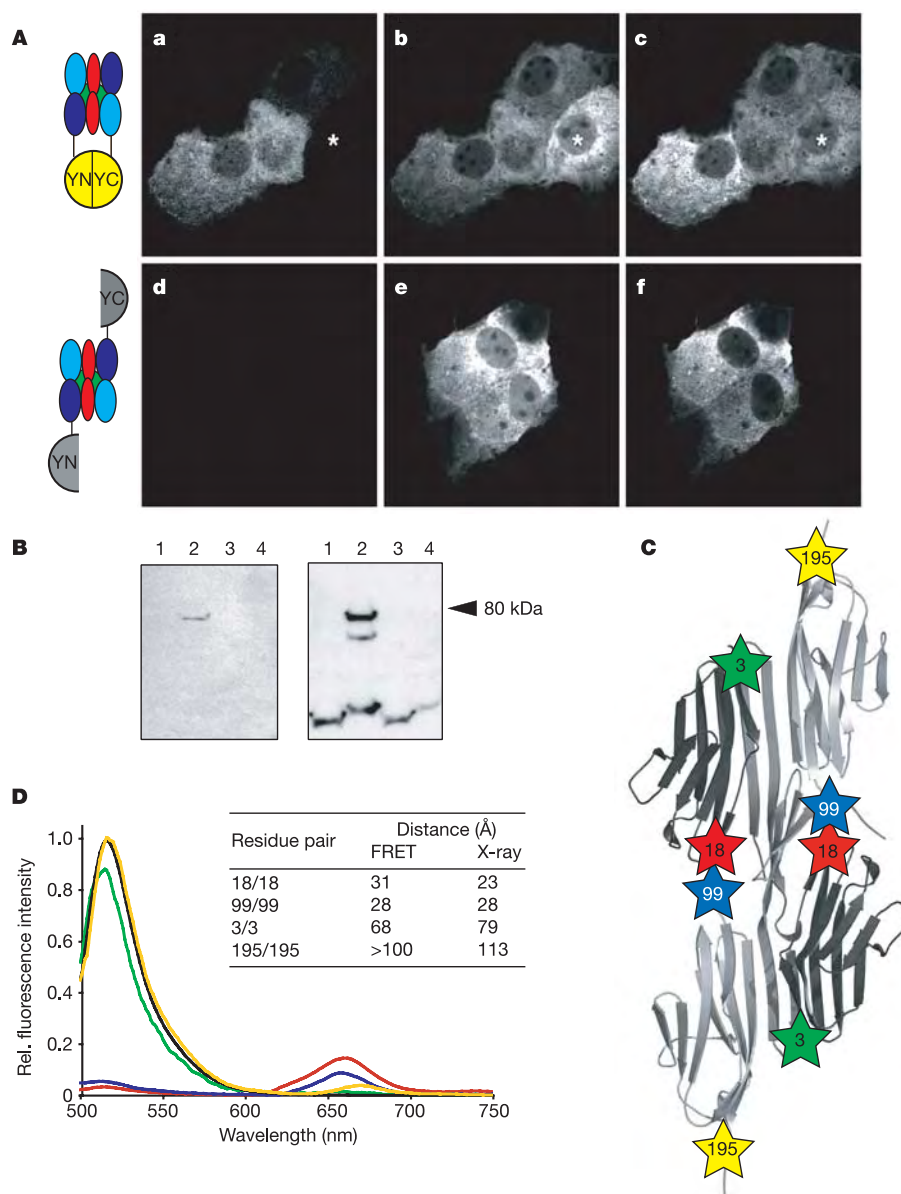


Figure 3 | Evidence for the formation of the palindromic titin–telethonin complex *in vivo* and *in vitro*. **A**, YFP reconstitution experiments in co-transfected COS cells. Top: N-terminal and C-terminal halves of YFP (YN, YC) were fused to opposite termini of the titin tandem immunoglobulin-like constructs YN–titin and titin–YC. Bottom: YN and YC were fused to the N terminus of the titin tandem immunoglobulin-like construct, yielding YN–titin and YC–titin. **a, d**, Intrinsic YFP fluorescence; **b, e**, detection of HA-epitope-tagged telethonin (residues 1–90), with an anti-HA antibody; **c, f**, YC detection, with a GFP-specific antibody. The asterisk in **c** indicates a COS cell expressing only one of the two split-GFP constructs, on the basis of the absence of intrinsic YFP fluorescence in **a**. **B**, Western blot analysis of whole COS cell extracts, using an anti-HA

antibody for telethonin detection (left) and a GFP antibody for the detection of split-GFP fusion constructs (right). COS cells were co-transfected with the HA-tagged titin-binding segment of telethonin (1–90; lanes 1 and 2), the C-terminal segment of telethonin (91–167; lanes 3 and 4) as well as either YN–titin and YC–titin (lanes 1 and 3) or YN–titin and titin–YC (lanes 2 and 4), respectively. **C, D**, FRET distance analysis of four titin residue pairs (identified by colour in **C**) from the titin–telethonin complex. The observed FRET distances (**D**, inset) can only be fitted with an antiparallel arrangement of the two titin molecules in the titin–telethonin complex. The spectrum of the donor-labelled titin(C195) mutant has been used as a reference (in black).

telethonin variants, in which the local hydrogen bond pattern of one of two the β -hairpin wings was disrupted by proline mutations, lost their capacity for binding the titin N terminus *in vitro* (Supplementary Fig. S3) and in COS-cell YFP complementation assays (Supplementary Fig. S4). In accordance with the molecular data, the same mutants were found to be unable to target correctly to the sarcomeric Z-disk when transfected into NRCs (Supplementary Fig. S2), indicating the abrogation of complex formation with sarcomeric titin. Taken together, the data indicate that the structural integrity of both pseudopalindromic telethonin wings might be critical for titin–telethonin assembly, regardless of whether telethonin assembles with the titin Z1Z2 domains only, as shown in COS cells, or via N-terminal titin within the sarcomeric Z-disk of intact myofibrils (Supplementary Table 1).

Our model implies that there is two-fold symmetry in the assembly of the N terminus of titin, in agreement with previous electron microscopy data¹⁷. Several arguments have been put forth indicating that titin–actin thin filaments might exist in a 2:1 ratio within the Z-disk^{6,17,18}. This stoichiometry could reconcile spacing considerations to match the 28/13 symmetry observed in actin thin filaments, allowing orthogonal α -actinin crosslinks at 195-Å intervals and satisfying the tetragonal lattice symmetry viewed along the filament axis as well as the estimated Z-repeat distances in titin in the order of 120 Å or less^{18,19}. An antiparallel titin–titin arrangement may be plausible because of the localization of the Z-links in the Z-disk centre as previously suggested^{7,18}. However, the titin–telethonin complex structure does not provide direct information about the origin of the two titin molecules that may belong to the same sarcomeric unit or to an adjacent sarcomere. The latter model, however, would inevitably lead to relative shifts of titin filaments outside the Z-disk areas in the range of hundreds of Ångströms, which would be in conflict with several imaging studies displaying titin as aligned filaments^{3,6,20}. The only reconcilable model therefore depicts the N termini of two titin strands as being derived from the same sarcomere.

Within the context of the Z-disk, our structure of the titin–telethonin complex provides an unexpected atomic model for the association of titin molecules at their very N termini, indicating that telethonin might act as a titin–titin crosslinker (Fig. 4). Its molecular architecture, along with evidence from binding and imaging data

both here and in previous papers^{3,4}, indicates that binding is very tight and may even be irreversible in the absence of signals that would weaken or degrade the interaction. The temporal delay in the proper localization of telethonin and its selective disappearance in neurogenic atrophy²¹ might indicate that telethonin turnover is regulated, either intrasterically or by as yet unknown alterations in the telethonin structure. In support of this, there is accumulating evidence indicating that the observed titin–telethonin assembly might interact with other protein components^{22–25} that generally seem to be more mobile than titin and telethonin^{7,26}.

The proposed titin–titin linker function of telethonin is analogous to that of the actin–titin linker of α -actinin⁶. However, in comparison with the α -actinin rod structure, the two terminal β -hairpin wing motifs in telethonin provide a much shorter linker, leading to a sandwich-type rather than to a rod-type linker model. In this complex, even though the shortest distance between the two titin N termini is only in the range of 4–15 Å, there are no direct specific titin–titin interactions. Thus, in structural terms, the function of telethonin is to tether the two titin N termini in close proximity to each other, unlike the α -actinin linker that provides a spacer of more than 200 Å (refs 6, 27). By unravelling the molecular basis of telethonin-mediated titin assembly, an overall picture is emerging on how the protein networks in the sarcomeric Z-disk may contribute to titin assembly and anchoring through at least two ligands (α -actinin and telethonin). This structure resists the mechanical forces generated in active muscle sarcomeres²⁸ and may feed back to the Z-disk stretch sensor machinery^{7,10}. Our data explain how some serious hereditary muscle diseases may be associated with the disruption of molecular interactions that connect and anchor sarcomeric filaments in the Z-disk by bridging mediators.

METHODS

Preparation of the titin–telethonin complex. A titin construct encoding domains Z1 and Z2 (1–196) and several telethonin variants comprising the full-length sequence (residues 1–167) or the N-terminal titin-binding region (1–90) were cloned, expressed and purified as described previously²⁹. In telethonin, Cys 8, Cys 15, Cys 38, Cys 57 and Cys 127 (1–167 only) were mutated into serine residues. Production of the seleno-L-methionine (SeMet)-incorporated telethonin is described in Supplementary Information.

Fluorescence imaging by *in vivo* complementation. Neonatal rat cardiomyocytes were prepared as described previously³⁰. For transfection assays, the pCMV-5 plasmid or the pEGFP plasmids (Clontech) were used. Telethonin was cloned bearing an N-terminal T7-tag sequence (MTGGQQMGR) or a carboxy-terminal green fluorescent protein (GFP) tag, because N-terminal GFP tags were found to act as dominant-negative proteins and to disrupt myofibrils. Transfection of cells was performed 1 day after plating with a standard liposome carrier system (Escort III) in accordance with the manufacturer's instructions (Sigma). At 24–48 h after transfection, cells were fixed in 4% paraformaldehyde/PBS for 10 min and stained with different antibodies as described previously³⁰. For the protein complementation experiments, titin(Z1Z2) complementary DNA was cloned by polymerase chain reaction and fused to either the N- or C-terminal region of YN(1–172) or YC(156–239) of YFP. COS-1 cells were co-transfected with haemagglutinin (HA)-tagged telethonin(1–90) or telethonin(91–163)–HA, together with either YN–titin(Z1Z2) and YC–titin(Z1Z2) or YN–titin(Z1Z2) and titin(Z1Z2)–YC. Cells were fixed 2 days after transfection and stained as described previously³⁰.

FRET analysis. Four different single cysteine-containing versions of the titin N terminus (Cys 3, Cys 18, Cys 99 and Cys 195) were used for labelling with Alexa488 (donor) and Cy5 (acceptor). For fluorescence measurements, the titin–telethonin complexes of the four mutants and wild-type titin(Z1Z2), as a negative reference, were mixed in the following molar ratios: protein:acceptor, 10:1; protein:acceptor, 1:40; protein:donor, 10:1; protein:donor, 1:40; protein:acceptor:donor, 1:50:1. The labelled probes were separated by gel filtration. To determine the concentration of each dye bound to the donor–acceptor sample adducts the absorbance was measured for $\lambda = 230$ –900 nm. The fluorescence spectra were scanned for $\lambda = 500$ –800 nm, with an excitation wavelength $\lambda_{ex} = 494$ nm. All experiments were performed in the dark. The energy transfer efficiency, E , of FRET was calculated as a function of the donor–acceptor distance (R_{AD}); $E = 1/(1 + (R_{AD}/R_0)^6)$, where R_0 is a DA-pair-specific constant, the Förster radius ($R_0 = 49$ Å) for the donor–acceptor pair used. Corrected fluorescence

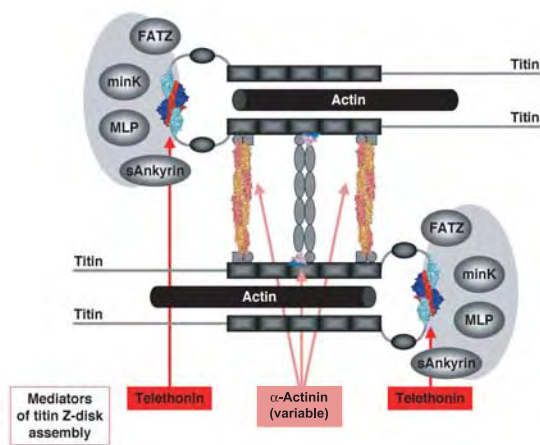


Figure 4 | Model outlining the involvement of the titin–telethonin complex in the architecture of the sarcomeric Z-disk. Titin filaments are assembled by a dual Z-disk bridging system, by α -actinin rods on a variable number of titin Z-repeats (three bridges are shown), and by telethonin by means of the N-terminal immunoglobulin domains Z1 and Z2. The titin N-terminus/telethonin complex forms a core that interacts with several ligands both inside and outside the sarcomeric Z-disk, including MLP, sAnkyrin, the β -subunit of the potassium channel (minK) and the γ -filamin/ABP-L, α -actinin and telethonin binding protein of the Z-disk (FATZ).

spectra were used to determine the quenching effect of the donor emission intensity when an acceptor was present within the FRET distance. The spectrum of the donor-labelled Z1Z2(C195) mutant (Z1Z2-D) was used as a reference. *E* values were determined directly from the fluorescence intensity: $E = 1 - I_{AD}/I_D$, where I_{AD} is the fluorescence intensity at 524 nm of the acceptor–donor labelled sample and I_D is the fluorescence intensity of Z1Z2-D at 524 nm.

X-ray structure determination. Crystals of the titin–telethonin(1–90) complex were grown by vapour diffusion, mixing equal volumes of about 15 mg ml^{−1} protein solution and mother liquid containing 5% w/v 8,000 kDa poly(ethylene glycol) (PEG 8000) and 100 mM sodium citrate buffer pH 4.45. Crystals were first grown in clusters of thin plates and were then used for macroseeding. In the seeding step, 6 mg ml^{−1} protein solution was mixed with 7.5% w/v PEG 35000, 100 mM sodium citrate buffer pH 4.45 and 200 mM Mg₂SO₄. X-ray data were collected on the tunable wiggler beamline BW6 (MPG-ASMB/DESY, Hamburg) and beamline X11 (EMBL/DESY, Hamburg). The X-ray structure was determined with the use of experimental phases from a selenomethionine version of the complex. Details of X-ray data acquisition, processing and structure determination are described in the legend to Supplementary Table S2.

Received 3 June; accepted 18 October 2005.

- Clark, K. A., McElhinny, A. S., Beckerle, M. C. & Gregorio, C. C. Striated muscle cytoarchitecture: an intricate web of form and function. *Annu. Rev. Cell Dev. Biol.* **18**, 637–706 (2002).
- Tskhovrebova, L. & Trinick, J. Titin: properties and family relationships. *Nature Rev. Mol. Cell Biol.* **4**, 679–689 (2003).
- Gregorio, C. C. *et al.* The NH₂ terminus of titin spans the Z-disc: its interaction with a novel 19-kD ligand (T-cap) is required for sarcomeric integrity. *J. Cell Biol.* **143**, 1013–1027 (1998).
- Mues, A., van der Ven, P. F., Young, P., Furst, D. O. & Gautel, M. Two immunoglobulin-like domains of the Z-disc portion of titin interact in a conformation-dependent way with telethonin. *FEBS Lett.* **428**, 111–114 (1998).
- Sorimachi, H. *et al.* Tissue-specific expression and α -actinin binding properties of the Z-disc titin: implications for the nature of vertebrate Z-discs. *J. Mol. Biol.* **270**, 688–695 (1997).
- Young, P., Ferguson, C., Banuelos, S. & Gautel, M. Molecular structure of the sarcomeric Z-disc: two types of titin interactions lead to an asymmetrical sorting of α -actinin. *EMBO J.* **17**, 1614–1624 (1998).
- Wang, J. *et al.* Dynamics of Z-band based proteins in developing skeletal muscle cells. *Cell Motil. Cytoskel.* **61**, 34–48 (2005).
- Pyle, W. G. & Solaro, R. J. At the crossroads of myocardial signaling: the role of Z-discs in intracellular signaling and cardiac function. *Circ. Res.* **94**, 296–305 (2004).
- Labeit, S. & Kolmerer, B. Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270**, 293–296 (1995).
- Knoll, R. *et al.* The cardiac mechanical stretch sensor machinery involves a Z disc complex that is defective in a subset of human dilated cardiomyopathy. *Cell* **111**, 943–955 (2002).
- Moreira, E. S. *et al.* Limb-girdle muscular dystrophy type 2G is caused by mutations in the gene encoding the sarcomeric protein telethonin. *Nature Genet.* **24**, 163–166 (2000).
- Hayashi, T. *et al.* Tcap gene mutations in hypertrophic cardiomyopathy and dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **44**, 2192–2201 (2004).
- Guddat, L. W. *et al.* Intramolecular signaling upon complexation. *FASEB J.* **9**, 101–106 (1995).
- Ramsland, P. A. & Farrugia, W. Crystal structures of human antibodies: a detailed and unfinished tapestry of immunoglobulin gene products. *J. Mol. Recognit.* **15**, 248–259 (2002).
- Remenyi, A., Scholer, H. R. & Wilmanns, M. Combinatorial control of gene expression. *Nature Struct. Mol. Biol.* **11**, 812–815 (2004).
- Fang, D. & Kerppola, T. K. Ubiquitin-mediated fluorescence complementation reveals that Jun ubiquitinated by Itch/AIP4 is localized to lysosomes. *Proc. Natl Acad. Sci. USA* **101**, 14782–14787 (2004).
- Liversage, A. D., Holmes, D., Knight, P. J., Tskhovrebova, L. & Trinick, J. Titin and the sarcomere symmetry paradox. *J. Mol. Biol.* **305**, 401–409 (2001).
- Luther, P. K. & Squire, J. M. Muscle Z-band ultrastructure: titin Z-repeats and Z-band periodicities do not match. *J. Mol. Biol.* **319**, 1157–1164 (2002).
- Atkinson, R. A. *et al.* Ca²⁺-independent binding of an EF-hand domain to a novel motif in the α -actinin–titin complex. *Nature Struct. Biol.* **8**, 853–857 (2001).
- Furst, D. O., Osborn, M., Nave, R. & Weber, K. The organization of titin filaments in the half-sarcomere revealed by monoclonal antibodies in immunoelectron microscopy: a map of ten nonrepetitive epitopes starting at the Z line extends close to the M line. *J. Cell Biol.* **106**, 1563–1572 (1988).
- Schroder, R. *et al.* Early and selective disappearance of telethonin protein from the sarcomere in neurogenic atrophy. *J. Muscle Res. Cell Motil.* **22**, 259–264 (2001).
- Kontogianni-Konstantopoulos, A. & Bloch, R. J. The hydrophilic domain of small ankyrin-1 interacts with the two N-terminal immunoglobulin domains of titin. *J. Biol. Chem.* **278**, 3985–3991 (2003).
- Faulkner, G., Lanfranchi, G. & Valle, G. Telethonin and other new proteins of the Z-disc of skeletal muscle. *IUBMB Life* **51**, 275–282 (2001).
- Frey, N. & Olson, E. N. Calsarcin-3, a novel skeletal muscle-specific member of the calsarcin family, interacts with multiple Z-disc proteins. *J. Biol. Chem.* **277**, 13998–14004 (2002).
- Furukawa, T. *et al.* Specific interaction of the potassium channel β -subunit mink with the sarcomeric protein T-cap suggests a T-tubule-myofibril linking system. *J. Mol. Biol.* **313**, 775–784 (2001).
- Mayans, O. *et al.* Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature* **395**, 863–869 (1998).
- Djinovic-Carugo, K., Young, P., Gautel, M. & Saraste, M. Structure of the α -actinin rod: molecular basis for cross-linking of actin filaments. *Cell* **98**, 537–546 (1999).
- Li, H. *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998–1002 (2002).
- Zou, P. *et al.* Solution scattering suggests cross-linking function of telethonin in the complex with titin. *J. Biol. Chem.* **278**, 2636–2644 (2003).
- Auerbach, D. *et al.* Different domains of the M-band protein myomesin are involved in myosin binding and M-band targeting. *Mol. Biol. Cell* **10**, 1297–1308 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Furst for the gift of antibodies; R. Kühnemuth for discussions on the FRET experiments; M. Forster for involvement in expression and purification tests; G. Burenkov for assistance during data collection at beamline BW6 at MPG-ASMB/DESY; E. Mandelkow and M. von Bergen for making the fluorimeter at MPG-ASMB/DESY available; and E. Ehler for the preparation of neonatal rat cardiomyocyte cultures. N.P. and S.L. were supported by the EU research and training network CAMKIN to M.W. and M.G., respectively. During involvement at the project, O.M. was supported by an EU Marie-Curie postdoctoral fellowship.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession number 1YA5. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.W. (wilmanns@embl-hamburg.de).

LETTERS

Structure of the Sec13/31 COPII coat cage

Scott M. Stagg^{1,2*}, Cemal Gürkan^{2*}, Douglas M. Fowler^{2,4}, Paul LaPointe², Ted R. Foss⁴, Clinton S. Potter^{1,2}, Bridget Carragher^{1,2} & William E. Balch^{2,3,5}

Endomembranes of eukaryotic cells are dynamic structures that are in continuous communication through the activity of specialized cellular machineries¹, such as the coat protein complex II (COPII), which mediates cargo export from the endoplasmic reticulum (ER)^{2,3}. COPII consists of the Sar1 GTPase, Sec23 and Sec24 (Sec23/24), where Sec23 is a Sar1-specific GTPase-activating protein and Sec24 functions in cargo selection, and Sec13 and Sec31 (Sec13/31), which has a structural role³. Whereas recent results have shown that Sec23/24 and Sec13/31 can self-assemble to form COPII cage-like particles⁴, we now show that Sec13/31 can self-assemble to form minimal cages in the absence of Sec23/24. We present a three-dimensional reconstruction of these Sec13/31 cages at 30 Å resolution using cryo-electron microscopy and single particle analysis. These results reveal a novel cuboctahedron geometry with the potential to form a flexible lattice and to generate a diverse range of containers. Our data are consistent with a model for COPII coat complex assembly in which Sec23/24 has a non-structural role as a multivalent ligand localizing the self-assembly of Sec13/31 to form a cage lattice driving ER cargo export.

Mammalian Sec13 and Sec31 genes were co-expressed in baculovirus-infected insect cells and the recombinant proteins were co-purified to homogeneity as judged by SDS–polyacrylamide gel electrophoresis (PAGE), immunoblotting and mass spectroscopy (LC-MS/MS) analyses⁵. Initial gel filtration chromatography (GFC) analysis of the Sec13/31 hetero-oligomers was consistent with that previously observed for yeast and mammalian Sec13/31 heterotetramers^{6–11}. However, after dialysis to improve sample solubility, GFC analysis (Fig. 1) revealed that mammalian Sec13/31 (Fig. 1, left inset) fractionated in the void volume (V_0) of the Superose 6 column. Considering the exclusion limit of this GFC media for globular proteins (40 MDa), one possibility was that Sec13/31 had aggregated after dialysis. However, dynamic light scattering (DLS) analysis suggested that the dialysed Sec13/31 sample comprised a relatively homogeneous population of particles with an average hydrodynamic radius (R_h) of ~400 Å (Fig. 1, right inset). Subsequent analysis of these Sec13/31 particles by analytical ultracentrifugation and GFC with online multi-angle light scattering (GFC–MALS) suggested a molecular mass distribution between 5.6 and 8.3 MDa, respectively. Electron microscopy analysis of the negatively stained samples revealed that purified Sec13/31 existed predominately as a collection of cage-like particles with diameters of 500–800 Å (Supplementary Fig. 1).

Because purified Sec13/31 forms a relatively homogeneous population of assemblies as judged by GFC, analytical ultracentrifugation, DLS, GFC–MALS and electron microscopy analyses of the negatively stained samples, we characterized the same samples using cryo-electron microscopy (cryo-EM) and single particle analysis. Images of the specimen preserved in vitreous ice (Fig. 1b) showed a

population of cage-like particles, most of which were symmetric and with an average diameter of ~600 Å. These dimensions are in good agreement with the size of COPII cages/vesicles observed *in vitro*^{4,12} and *in vivo*^{13,14}.

For single particle analysis, a total of 9,777 individual cage particles were selected from a set of 516 defocus pairs of micrographs. Particles were first subjected to a reference-free alignment algorithm as implemented in the EMAN package¹⁵ to generate averages with an improved signal-to-noise ratio. Of the 104 resulting class averages, ten that showed the best signal-to-noise ratio and symmetry (Supplementary Fig. 2a) were used as reference images in a multi-reference alignment procedure. The resulting class averages exhibited two-fold, three-fold and four-fold symmetry (Supplementary Fig. 2b) and geometry consistent with that of a cuboctahedron (Supplementary Fig. 2c).

Cuboctahedrons are roughly spherical polyhedrons with 24 edges and 12 vertices, of which 8 are triangles and 6 are squares, and exhibit 4 3 2 or octahedral symmetry. The geometry of a cuboctahedron is defined by the intersection of four edges at each vertex, in contrast to clathrin geometries, which are defined by vertices formed from only three edges. For a cuboctahedron, the four-fold rotational axes of symmetry run down the middle of the square faces, the three-fold rotational axes run through the middle of the triangular faces, and the two-fold rotational axes run through the vertices (Supplementary Fig. 2c).

Single particle methods were used to reconstruct a three-dimensional electron microscopy density map of the mammalian Sec13/31 cages. Using a simple cuboctahedron constructed with continuous density for the edges as an initial model, the cage structure was refined to a resolution of 30 Å. There is excellent agreement between projections of the final model and the individual raw particle images as well as the class averages (Fig. 2). Initial model bias was explored using a variety of other starting structures (see Methods), all of which either converged to a cuboctahedron or failed to converge to a consistent structure. The reconstructed cage structure (Fig. 3a and Supplementary Video 1) has a diameter of 600 Å along its longest diagonal, the length of an edge is 300 Å, and the width of an edge is 40 Å. The molecular mass is estimated to be 5.4–9.6 MDa based on the volume of electron density corresponding to the reconstructed Sec13/31 cage (Fig. 3a). This is consistent with molecular mass measurements obtained from analytical ultracentrifugation and GFC–MALS analyses.

The asymmetric unit (ASU) of the cuboctahedral cage—the smallest unit that can be repeated to generate the full structure—is shown in Fig. 3b. The ASU consists of two roughly spherical lobes of density (one small and one large) at either end (Fig. 3b, regions 1, 2, 5 and 6) connected by a continuous curving stretch of density with a diameter of 40 Å (Fig. 3b, regions 3 and 4). Although the lobes at either end are not identical, they appear to be related to each other by

¹National Resource for Automated Molecular Microscopy, Departments of ²Cell, ³Molecular Biology and ⁴Chemistry, and ⁵The Institute for Childhood and Neglected Disease, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

*These authors contributed equally to this work.

a 180° rotation around the centre of the density connecting the two ends. The fact that this two-fold symmetry is not imposed during the reconstruction implies that the ASU is a dimer. It is interesting to note that the centre of symmetry of the ASU is not in the centre of the edge of the cuboctahedron. Overall, the dimensions of the edges of the mammalian Sec13/31 cage are in close agreement with those of the proposed heterotetramer structure of the yeast Sec13/31 complex observed in negative stain⁷ and quick-freeze/deep-etch rotary shadowing scanning electron microscopy¹³. Therefore, we propose that the 24 off-centre, dimeric ASUs comprising the Sec13/31 cage (Fig. 3a) correspond to 24 Sec13/31 heterotetramers (Fig. 4a).

Our results now provide a 30 Å resolution map of the structure of the cage generated by the self-assembling properties of Sec13/31 hetero-oligomers. Our ability to reconstitute Sec13/31 self-assembly *in vitro* contrasts with previous studies using yeast and mammalian Sec13/31 purified from either recombinant or native sources^{6,8,9,16,17}. DLS analysis of unassembled Sec13/31 suggests an average particle R_h of 106 ± 27 Å, which is consistent with the heterotetramers observed

in earlier studies⁴. After dialysis⁵, DLS analysis of the Sec13/31 samples reveals a relatively homogeneous particle population with an average R_h of 407 ± 9.4 Å. This suggests that Sec13/31 heterotetramers self-assemble to form a cage-like structure. Together, analytical ultracentrifugation, GFC–MALS and electron microscopy density analyses suggested a molecular mass distribution of 5.4 to 9.6 MDa for the assembled cages. These results are consistent with our proposed model (Fig. 4a) in which 24 Sec13/31 hetero-oligomers come together to make a cuboctahedral cage with a calculated mass of 8.14 MDa.

The positions occupied by Sec13 and Sec31 in the cage remain to be determined. From a structural perspective, Sec13 contains WD40 repeat motifs that are implicated in protein–protein interactions¹⁸. Biochemical and computational analyses indicate that the Sec13 structure may comprise a single domain, β -propeller fold with six blades^{19,20}. Although Sec31 contains WD40 motifs at its amino-terminal domain that are proposed to form a β -propeller with seven blades^{9,19}, the rest of the protein is predicted to have an α -solenoid fold, a distinct domain arrangement also shared by clathrin^{19,21,22}. Given the size and orientation of yeast Sec13/31 observed by electron microscopy analysis of negatively stained samples⁷, the heterotetramer model proposed previously (Sec31/Sec13–Sec13/Sec31) could constitute the ASU of the Sec13/31 cage. In this view, two Sec13 proteins would form part of the continuous density in the centre of the ASU (3 and 4 in Fig. 3b), but cannot be resolved as distinct entities at the present resolution. Such a model would suggest that Sec13 dimerization is critical for cross-bridging the two halves of the edge.

An alternative model that is more consistent with biochemical data and now takes into account the new Sec13/31 cage structure is that Sec13 forms the vertices of the cuboctahedron. This would suggest that the Sec13/31 heterotetramer is arranged as Sec13/Sec31–Sec31/

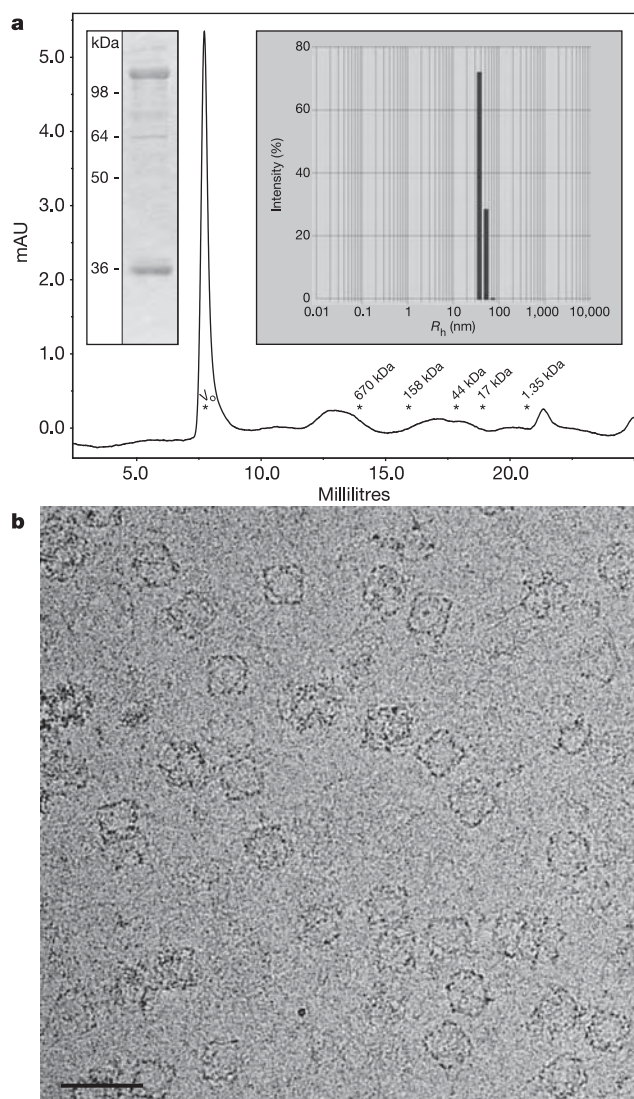


Figure 1 | Initial analyses of the self-assembled Sec13/31 cages. **a**, Gel filtration chromatography of the dialysed Sec13/31 sample. Coomassie-stained SDS–PAGE of assembled Sec13/31 particles (left inset). DLS analysis showing the distribution of the hydrodynamic radii (R_h) of the self-assembled Sec13/31 particles (right inset). **b**, A typical far-from-focus micrograph of the Sec13/31 cages preserved in vitreous ice. Scale bar, 1,000 Å. mAU, milliabsorbance units.

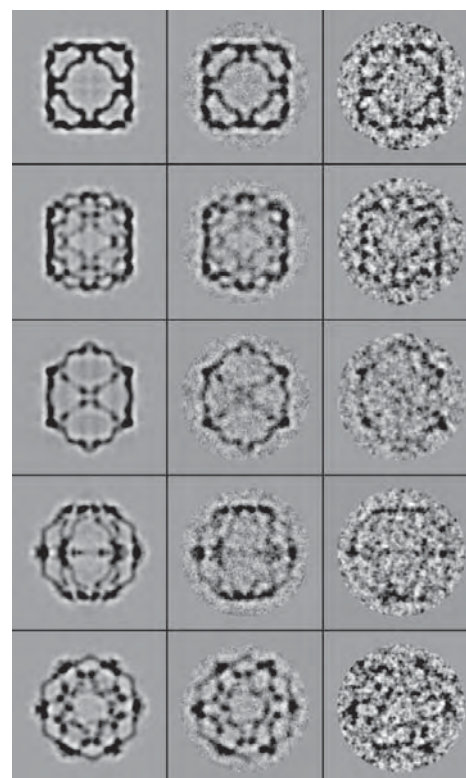


Figure 2 | Comparison of the refined Sec13/31 cage structure to the raw data. The left column shows projections of the refined map; the middle column shows corresponding class averages; and the right column shows corresponding individual raw particle images.

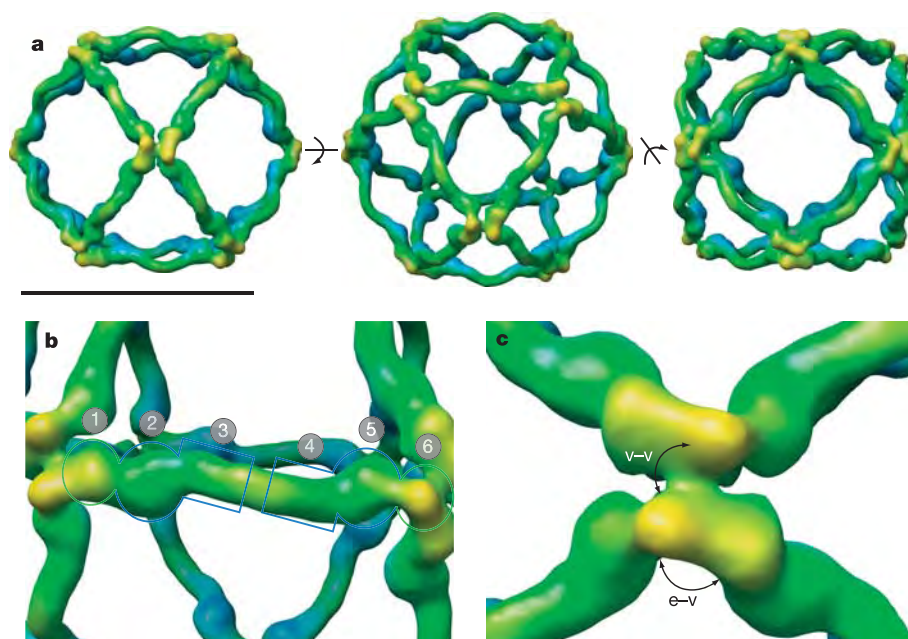


Figure 3 | 30 Å resolution map of the Sec13/31 cage. **a**, Views of the cage along its two-fold (left panel), three-fold (centre panel) and four-fold (right panel) axes of symmetry. The surface of the cage is coloured from blue (nearest) to yellow (farthest) according to its distance from the centre of the volume. Scale bar, 500 Å. **b**, The asymmetric unit and vertex of the

cuboctahedral Sec13/31 lattice. Structural features are numbered 1–6. The proposed Sec13/31 heterotetramer arrangement is indicated by the green lines for Sec13 and blue lines for Sec31. **c**, Close-up of a cage vertex. Vertex–vertex (v–v) and edge–vertex (e–v) interactions are indicated.

Sec13 and corresponds to the ASU that constitutes the edges of the cuboctahedron as shown in the two-dimensional lattice (Fig. 4a). In this view, regions 1 and 6 (Fig. 3b) that form the cuboctahedral vertices in the reconstructed cage contain Sec13. The Sec13 subunits would interact with each other at the vertices of the cage in two unique ways: via edge–vertex contacts and vertex–vertex contacts (Fig. 3c). This is consistent with the recent observation that Sec13 alone in solution can form dimers and tetramers⁷. Moreover, the larger globular domains (2 and 5 in Fig. 3b) would correspond to the

predicted β -propeller fold comprising the seven WD40 motif repeats/blades of the Sec31 N-terminal domain^{9,19}, as opposed to the smaller globular domains (1 and 6 in Fig. 3b) that would correspond to the predicted β -propeller fold comprising the six WD40 motif repeats/blades of the entire Sec13 subunit^{19,20}. Considering that Sec13 interacts with the N-terminal WD40 repeat domain of Sec31 (refs 9, 23), this new model would place the N-terminal domain of Sec31 near the vertex of the cuboctahedral cage. It follows that Sec31 dimerization at the centre of the ASU would be critical for cross-bridging the two halves of the cuboctahedral edge. In support of this view, after dissociation of Sec13 from the Sec13/31 heterotetramer by 2 M urea, GFC reveals that Sec31 still fractionates as a 500–700 kDa species⁶, suggesting that loss of Sec13 does not compromise Sec31 dimerization. Given the biochemical evidence that the Sec24 subunit interacts with the middle region of Sec31 (ref. 23) and the Sec23 subunit interacts with the proline-rich region towards the Sec31 carboxy terminus⁹, Sec23 is expected to bind near the Sec31–Sec31 dimer interface, whereas Sec24 should bind towards the ends of the ASU (Fig. 4b).

Clathrin can also self-assemble *in vitro* to form empty cages lacking the adaptor components and cargo, all of which comprise the clathrin coat^{21,22,24}. These are strikingly different from the Sec13/31 cage. Although three edges intersect to form the vertices in clathrin geometries, Sec13/31 self-assembles into a cuboctahedron where four edges intersect to form the vertices. Additionally, each edge of the clathrin cage is formed from four overlapping clathrin heavy chains with a diameter of 100 Å (refs 21, 22, 24). In contrast, we propose that each edge of the Sec13/31 cage is comprised of a Sec13/31 heterotetramer with a diameter of 40 Å. Thus, the edges of the Sec13/31 cuboctahedron are unlikely to be as extensively interdigitated as those of clathrin. These results suggest that different biochemical and structural solutions have evolved for constructing self-assembling polymers that can generate transport containers.

COPII vesicles must be capable of expanding to accommodate cargo of varying sizes^{25,26}. Recent studies have demonstrated the presence of highly pleomorphic coated structures at ER exit sites that house large cargoes such as chylomicron particles²⁵ and pro-collagen

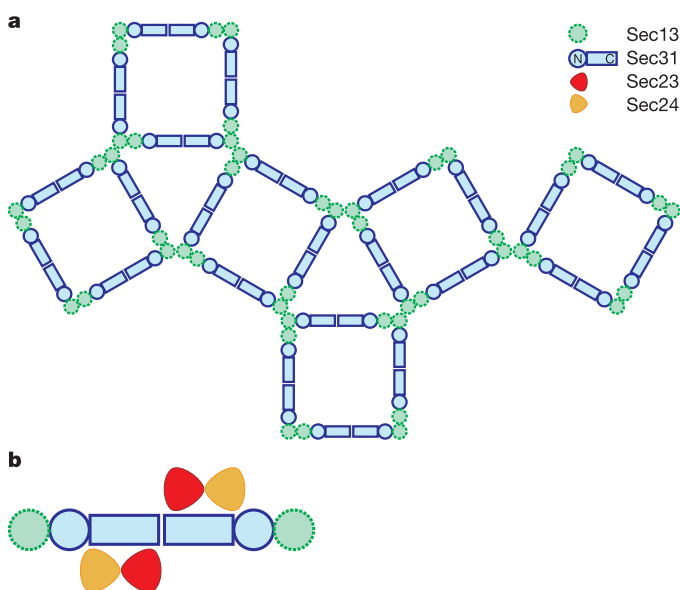


Figure 4 | Orientation of the Sec13/31 heterotetramer in the self-assembled cuboctahedral cage. **a**, A flattened version of the cuboctahedral cage structure illustrates the proposed orientation of the Sec13/31 heterotetramer at the edges and vertices. **b**, Expected locations of Sec23 and Sec24 subunits on the Sec13/31 heterotetramer.

polymers²⁶ that are too large to be incorporated into standard 600-Å sized COPII vesicles. As with the clathrin cage²², the Sec13/31 cage could solve this problem in part by forming lattices with different geometries. There are four other symmetrical polyhedrons that have four identical edges that meet at a vertex. In order of increasing size, these are the octahedron, the small rhombicuboctahedron, the icosidodecahedron and the small rhombicosidodecahedron (Supplementary Fig. 3). Thus, Sec13/31 may be able to accommodate cargoes with a wide range of sizes by forming cages with different numbers of edges. These conclusions are consistent with observations of less abundant but larger cage structures in our cryo-EM images (data not shown). The novel cuboctahedron geometry of the Sec13/31 cages may also provide the COPII coat structure with the ability to flex to accommodate cargo of different shapes during the recruitment of the adaptor components such as Sec23/24. We suggest that the unique structural properties of the Sec13/31 cage may enable it to form flexible polyhedrons of increasingly larger geometries that are capable of incorporating large, oddly shaped cargo.

Our results now demonstrate that the function of Sec13/31 is analogous to that of clathrin, which self-assembles to form a cage independent of its adaptor proteins^{21,22,24}. This is in contrast with a recent study that suggested that the Sec23/24 adaptor is required for the self-assembly of a minimal COPII cage⁴. Our data are now consistent with a model for COPII coat formation where Sec23/24, like the clathrin adaptor proteins, coordinates cargo selection with the self-assembly of the Sec13/31 cage to promote budding from the ER. The discovery of the self-assembling properties of Sec13/31 to generate a cage structure provides a new focus for elucidating the biological mechanisms of cargo selection, concentration and budding for transport of nearly one-third of all proteins encoded by the eukaryotic genome.

METHODS

Recombinant production and purification. Recombinant expression and purification to homogeneity of the mammalian Sec13/31 hetero-oligomers in baculovirus-infected insect cells is as described⁵. Briefly, human *SEC13R* (also known as *SEC13L1*) and *SEC31L1* genes (GenBank accession numbers NM_183352 and NM_014933, respectively) were cloned into the pFastBac DUAL expression vector (catalogue number 10359-016) and recombinant co-expression was carried out in Tn5 insect cells (catalogue number B85502) (Invitrogen). Recombinant SEC13R has an N-terminal hexa-histidine tag for initial purification of Sec13/31 by immobilized metal affinity chromatography. This is followed by anion exchange chromatography and dialysis of the Sec13/31-rich pool in low salt buffer (20 mM Tris-Cl, pH 7.5, 300 mM NaCl, 1 mM MgOAc, 10 mM dithiothreitol (DTT)) against a high salt buffer (25 mM HEPES, pH 7.5, 700 mM KOAc, 1 mM MgOAc, 1 mM DTT) before storage at -80 °C.

Dynamic light scattering. DLS analyses were carried out at 25 °C using a DynaPro-MS/MSTC controlled by Dynamics V5.25.44 software (Wyatt Technology). Fifteen microlitres of purified Sec13/31 (~0.1 mg ml⁻¹) in the high salt buffer were transferred to a quartz cuvette and allowed to equilibrate at 25 °C before taking on average ten readings with an acquisition time of 10 s each. Alternatively, DLS analysis was carried out using a DAWN EOS/WyattQELS system equipped with a Peltier temperature controller (Wyatt Technology). At least 15 autocorrelation function measurements were taken at 25 °C in 30 µl quartz microcuvettes and single exponential fits were performed to derive R_h and R_h numerical averages were taken using the ASTRA v4 software.

Electron microscopy. Electron microscopy of uranyl acetate-stained samples was carried out using a FEI Tecnai F20 instrument (Philips Electron Optics) equipped with a Gatan UltraScan 4000 digital camera. For cryo-EM analysis, self-assembled Sec13/31 cages were dialysed into 50 mM MES pH 6.5, 225 mM KOAc, 1 mM MgCl₂ to minimize background during imaging. DLS analysis confirmed that this step does not affect the self-assembled Sec13/31 cages. Sec13/31 cages were preserved in vitreous ice by placing 4 µl of the dialysed sample onto Quantifoil 2/2 400 mesh grids (Quantifoil Micro Tools) that had been plasma cleaned for 30 s using a Fischione model 1020 plasma cleaner (Fischione Instruments). The grids were blotted and plunged into liquid ethane using an FEI Vitrobot. Two data sets (A and B) were collected on a FEI Tecnai F20 electron microscope operated at 120 and 200 keV, respectively. Magnification (×50,000)

exposure pairs with defoci ranging from 1.0 to 4.5 µm under focus were collected under low dose conditions using the Leginon automated electron microscopy package²⁷. A total of 138 and 516 exposure pairs were acquired for the first and second data sets, respectively.

Single particle reconstruction. Individual particles were selected from the micrographs using the Selexon automated particle picking package²⁸. The automated selections were then manually inspected, eliminating bad particles and selecting missed particles using the boxer program from the EMAN package¹⁵. A total of 2,235 and 9,777 particles were selected from the first (A) and second (B) data sets, respectively. The contrast transfer function (CTF) for each micrograph was estimated using the ACE automated CTF estimation package²⁹. CTF parameters for the far-from-focus images were estimated, and the phases were flipped for the near-to-focus images using defocus estimated from the far-from-focus images with appropriate adjustments for the difference in their defoci.

Single particle reconstruction refinement was performed using the EMAN package. The starting model for the refinement was a cuboctahedron of approximately the same size as the cages and with continuous density for the edges. Using particles from data set A, the model was refined for six iterations starting with an angular increment of seven going down to an angular increment of four. Octahedral symmetry was imposed during the refinement. The resolution for the reconstruction was calculated using the 0.5 Fourier shell correlation (FSC) criteria to be 43 Å. This model was used as an initial model in a second refinement using data set B. The model was further refined for seven iterations with increasingly stringent conditions for inclusion of a particle. The resolution for the final model was 30 Å according to the 0.5 FSC criteria.

To explore the effect of initial model bias, both a simple cube and an octahedron were used as initial starting models in refinements. Starting with the cube, the structure converged back to a cuboctahedron after six iterations. Using the model octahedron in the standard orientation for EMAN, the refined structure resulted in projections that did not match the class averages. When the starting octahedron was rotated 45° along the z axis, which preserved one of the four-fold axes of symmetry but none of the other axes, the model once again converged back to a cuboctahedron within five iterations of refinement.

Received 22 June; accepted 18 October 2005.

- Gurkan, C. et al. Large-scale profiling of Rab GTPase trafficking networks: The membrane. *Mol. Biol. Cell* **16**, 3847–3864 (2005).
- Bonifacino, J. S. & Glick, B. S. The mechanisms of vesicle budding and fusion. *Cell* **116**, 153–166 (2004).
- Lee, M. C., Miller, E. A., Goldberg, J., Orci, L. & Schekman, R. Bi-directional protein transport between the ER and Golgi. *Annu. Rev. Cell Dev. Biol.* **20**, 87–123 (2004).
- Antony, B., Gounon, P., Schekman, R. & Orci, L. Self-assembly of minimal COPII cages. *EMBO Rep.* **4**, 419–424 (2003).
- Gurkan, C. & Balch, W. E. Recombinant production in baculovirus-infected insect cells and purification of the mammalian Sec13/Sec31 complex. *Methods Enzymol.* **404**, 58–66 (2005).
- Salama, N. R., Chuang, J. S. & Schekman, R. W. Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum. *Mol. Biol. Cell* **8**, 205–217 (1997).
- Lederkremer, G. Z. et al. Structure of the Sec23p/24p and Sec13p/31p complexes of COPII. *Proc. Natl Acad. Sci. USA* **98**, 10704–10709 (2001).
- Kim, J., Hamamoto, S., Ravazzola, M., Orci, L. & Schekman, R. Uncoupled packaging of amyloid precursor protein and presenilin 1 into COPII vesicles. *J. Biol. Chem.* **280**, 7758–7768 (2005).
- Shugrue, C. A. et al. Identification of the putative mammalian orthologue of Sec31P, a component of the COPII coat. *J. Cell Sci.* **112**, 4547–4556 (1999).
- Salama, N. R., Yeung, T. & Schekman, R. W. The Sec13p complex and reconstitution of vesicle budding from the ER with purified cytosolic proteins. *EMBO J.* **12**, 4073–4082 (1993).
- Tang, B. L. et al. Mammalian homologues of yeast sec31p. An ubiquitously expressed form is localized to endoplasmic reticulum (ER) exit sites and is essential for ER-Golgi transport. *J. Biol. Chem.* **275**, 13597–13604 (2000).
- Matsuoka, K. et al. COPII-coated vesicle formation reconstituted with purified coat proteins and chemically defined liposomes. *Cell* **93**, 263–275 (1998).
- Matsuoka, K., Schekman, R., Orci, L. & Heuser, J. E. Surface structure of the COPII-coated vesicle. *Proc. Natl Acad. Sci. USA* **98**, 13705–13709 (2001).
- Aridor, M., Bannykh, S. I., Rowe, T. & Balch, W. E. Cargo can modulate COPII vesicle formation from the endoplasmic reticulum. *J. Biol. Chem.* **274**, 4389–4399 (1999).
- Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
- Belden, W. J. & Barlowe, C. Purification of functional Sec13p-Sec31p complex, a subunit of COPII coat. *Methods Enzymol.* **329**, 438–443 (2001).
- Shimoni, Y. & Schekman, R. Vesicle budding from endoplasmic reticulum. *Methods Enzymol.* **351**, 258–278 (2002).

18. Pryer, N. K., Salama, N. R., Schekman, R. & Kaiser, C. A. Cytosolic Sec13p complex is required for vesicle formation from the endoplasmic reticulum *in vitro*. *J. Cell Biol.* **120**, 865–875 (1993).
19. Devos, D. *et al.* Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* **2**, e380 (2004).
20. Saxena, K. *et al.* Analysis of the physical properties and molecular modeling of Sec13: A WD repeat protein involved in vesicular traffic. *Biochemistry* **35**, 15215–15221 (1996).
21. Musacchio, A. *et al.* Functional organization of clathrin in coats: combining electron cryomicroscopy and X-ray crystallography. *Mol. Cell* **3**, 761–770 (1999).
22. Fotin, A. *et al.* Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* **432**, 573–579 (2004).
23. Shaywitz, D. A., Espenshade, P. J., Gimeno, R. E. & Kaiser, C. A. COPII subunit interactions in the assembly of the vesicle coat. *J. Biol. Chem.* **272**, 25413–25416 (1997).
24. Smith, C. J., Grigorieff, N. & Pearse, B. M. Clathrin coats at 21 Å resolution: a cellular assembly designed to recycle multiple membrane receptors. *EMBO J.* **17**, 4943–4953 (1998).
25. Jones, B. *et al.* Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders. *Nature Genet.* **34**, 29–31 (2003).
26. Mironov, A. A. *et al.* ER-to-Golgi carriers arise through direct en bloc protrusion and multistage maturation of specialized ER exit domains. *Dev. Cell* **5**, 583–594 (2003).
27. Suloway, C. *et al.* Automated molecular microscopy: The new Legimon system. *J. Struct. Biol.* **151**, 41–60 (2005).
28. Zhu, Y., Carragher, B. & Potter, C. S. *Improving Template Matching Based Particle Selection for Cryo-electron Microscopy* (IEEE ISBIO4 Conference, Arlington, Virginia, 2004).
29. Mallick, S. P., Carragher, B., Potter, C. S. & Kriegman, D. J. ACE: Automated CTF Estimation. *Ultramicroscopy* **104**, 8–29 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements These studies are supported by NIH grants to W.E.B., C.S.P. and B.C. S.M.S. is a recipient of an NIH Postdoctoral Fellowship. C.G. is a recipient of a Cystic Fibrosis Foundation Postdoctoral Research Fellowship. We thank G. Cantin and J. Venable for LC-MS/MS analysis, and J. T. Weissman, K. Straley and S. J. Lloyd for setting up or helping with recombinant expression of the human Sec13/31 complex in baculovirus-infected insect cells. This is TSRI manuscript number 17546-CB.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.C. (bcarr@scripps.edu) or W.E.B. (webalch@scripps.edu).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

The best-laid plans

A career in science requires planning — from drawing up grant proposals to seeking tenure. But, as for most things in life, events rarely follow the path laid out. This has readily been demonstrated over the past two years by the eight PhD students who have detailed their experiences in our Graduate Journal column.

Most PhD students intend to finish their studies in a set number of years, emerging with tangible results and, ideally, solid job opportunities. But is that usually the case? To find out, we have invited the four students who launched our Graduate Journal in 2004 to return this month for an alumni special.

The first to report back is Sidney Omelon, a PhD candidate at the University of Toronto, who is finding that writing a thesis can take longer and present more challenges than initially predicted (see page 240).

And the class of 2005, who wrapped up their contributions last month, have already shown how unpredictable life can be. Jason Underwood, for example, expected to complete his PhD last year at the University of California, Los Angeles, then settle down to a local job and

life with his wife. By February, his wife had left him and his carefully laid plans had to be rethought (see *Nature* **433**, 782; 2005).

Why should we make a note of such twists and turns? Because one running theme in *Naturejobs* is the need to evaluate risks. This is useful at all stages and skill levels of the scientific process.

But another reason is that *Naturejobs* is this year running a competition to find its next crop of PhD journal-keepers. Students interested in sharing their experiences should visit www.nature.com/naturejobs/magazine/competition.html — but be quick as entries must be received by 20 January. We can make no guarantees of what will come of this process, only that, if the past two years are anything to go by, it is bound to be interesting.



Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

Natureevents: Sille Opstrup (4994)
UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
Germany/Austria/The Netherlands:
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/
The Netherlands:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 75 Varick Street,
 9th Floor, New York,

NY 10013-1917
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

MOVERS

Miodrag Stojkovic, deputy director of regenerative medicine, Prince Felipe Research Centre, Valencia, Spain



2004–05: Deputy director, Centre for Stem Cell Biology and Developmental Genetics, University of Newcastle upon Tyne, UK

2003–05: Reader in embryology and stem-cell biology, University of Newcastle upon Tyne, UK

1998–2002: Head of an *in vitro* fertilization laboratory, Ludwig-Maximilians University, Munich, Germany

Born in Leskovac in former Yugoslavia, Miodrag Stojkovic has in the past 15 years moved to Germany, then to Britain and has recently arrived in Spain. "That's it for me — no more languages after this one," he jokes. Indeed, his ability to adapt to new countries has served him well as he takes up his post at Spain's newest biomedical research centre.

After receiving his degree in veterinary medicine at the University of Belgrade, Stojkovic left his native country for Germany in 1991, just before the Balkan conflict started. "I had a feeling the war was coming," he says. For two years he worked as a nurse at the University Hospital of Hamburg before moving to Munich to study veterinary medicine again. Stojkovic had to repeat four semesters of coursework, pass additional exams and attend a language school to get his degree recognized in Germany. He began a PhD programme at Ludwig-Maximilians University in Munich and worked as a nurse during weekends.

After receiving his doctorate, he stayed on and soon became the head of an animal *in vitro* fertilization lab. He considers his time in Munich to be the most important period of his career. "It was here that I learned all the basic techniques that I am now using for my work on human embryos and human embryonic stem cells," he says.

Wanting to shift his work from animal embryos to human ones, Stojkovic soon realized that Germany's law restricting work on human embryonic stem cells would force him to move. In 2002 he went to Britain, where he joined the Newcastle Fertility Centre and the Institute of Human Genetics at the Centre for Life. Six months later he derived Britain's first fully characterized human embryonic stem-cell line. He then became a reader at the University of Newcastle upon Tyne, and three years later he created Europe's first cloned human embryos. Stojkovic also became deputy director of the university's stem-cell biology centre.

Stojkovic first got to know the Prince Felipe Research Centre when he was invited to give a lecture there. He was immediately impressed by the atmosphere and the commitment of its researchers. Realizing that Spain would offer him better opportunities, he decided to move. His belief in the therapeutic potential of stem cells motivates him. "If you believe in something, it will give you power and energy, no matter what political or religious obstacles you might come across," he says.

Siëlle Gramser

MENTORS & PROTÉGÉS

Teaching in depth

My adviser John Krommes, a theoretical plasma physicist at Princeton University, is an active researcher and outstanding teacher. He takes as much time as needed with students and watches out for our career development. That is why he received a Graduate Mentoring Award last year from Princeton.

Good mentoring begins with good teaching. In plasma physics, it is easy for students and professors to neglect the fundamentals and work from established results. That does not happen with John. He teaches from first principles. Students in his class are guided by more than 300 pages of handouts that he updates every year.

John makes plenty of time for his students. When Jill Foley, one of my fellow students, was preparing for her general exams, she asked John for a brief meeting to go through some questions. They spent eight hours reviewing the most challenging topics from two years of plasma-physics courses. "He didn't just give me the answers but led me to them so that I knew how to take that path again," says Jill.

John nurtures young scientists by really listening to our ideas. For example, we recently worked on the creation of fake data to test a data-analysis technique our team is using. John wanted to focus on the fundamental

plasma physics required to simulate data that more closely resemble the experimental data. But my thoughts on this problem led me to new understandings of the analysis technique itself. Instead of pursuing the topic of greatest interest to him, John took about five hours to understand my questions and ideas and to discuss them with our experimental partners. After much thought, he sent me an 800-word e-mail guiding my next steps.

John also contributes to our future as researchers. For example, even though both John and I are primarily analytic theorists, he has suggested that I invest significant project time acquiring skills in numerical simulation. This will probably not be essential for my thesis but may be important for my employment prospects.

The demands of pursuing a thesis and the all-too-frequent struggles with demanding, abusive or absentee advisers seem to cause a lot of soul-searching among graduate students about their career paths. But I know that after graduation, I will be seeking a research job as a plasma physicist. John Krommes deserves a great deal of credit for that confidence.

Timothy Stoltzfus-Dueck, a graduate student at Princeton University, drew on the experiences of several fellow students to compile this tribute.

ALUMNUS JOURNAL

Writing up

I was a Graduate Journal writer for *Naturejobs* in 2004 and have spent the past year finishing my PhD. The physical and mental toll of writing it up took me by surprise. I realized that not only my results, but my experimental designs, the criteria I used to review the literature, my interpretations and my understanding of first principles would all be under scrutiny. As I strove to explain my results, I gaped at the seemingly infinite magnitude of what I do not know.

As much as it was painful and humbling, I also found writing was a reflective and rewarding experience. It was a luxury to be able to focus on all aspects of my work at once, connecting the results from different experiments and finally bringing them together into an illustrative story.

I am in awe of the number of people who helped me produce this thesis. Listing the names of technicians, fellow students and a supervisor (who lived up to his job description) on an acknowledgements page seems to be paltry payback. I received much helpful advice, including one tip from a fellow student: "The most important thing to get out of graduate school is yourself."

Provided that my defences go well, I will be out by early this year. My time, tuition and stress are not adequately represented by those 201 pages. But I hope that what I do in the future will be a credit to the talented people who helped me gain the skills and knowledge that I take away with me.

Sidney Omelon will soon graduate from the University of Toronto, Canada.

Printcrime

Copy this story.

Cory Doctorow

The coppers smashed my father's printer when I was eight. I remember the hot, cling-film-in-a-microwave smell of it, and Da's look of ferocious concentration as he filled it with fresh goop, and the warm, fresh-baked feel of the objects that came out of it.

The coppers came through the door with truncheons swinging, one of them reciting the terms of the warrant through a bullhorn. One of Da's customers had shopped him. The ipolice paid in high-grade pharmaceuticals — performance enhancers, memory supplements, metabolic boosters. The kind of things that cost a fortune over the counter; the kind of things you could print at home, if you didn't mind the risk of having your kitchen filled with a sudden crush of big, beefy bodies, hard truncheons whistling through the air, smashing anyone and anything that got in the way.

They destroyed grandma's trunk, the one she'd brought from the old country. They smashed our little refrigerator and

the purifier unit over the window. My tweetybird escaped death by hiding in a corner of his cage as a big, booted foot crushed most of it into a sad tangle of printer-wire.

Da. What they did to him. When he was done, he looked like he'd been brawling with an entire rugby side. They brought him out the door and let the newsies get a good look at him as they tossed him in the car. All the while a spokesman told the world that my Da's organized-crime bootlegging operation had been responsible for at least 20 million in contraband, and that my Da, the desperate villain, had resisted arrest.

I saw it all from my phone, in the remains of the sitting room, watching it on the screen and wondering how, just how anyone could look at our little flat and our terrible, manky estate and mistake it for the home of an organized crime kingpin. They took the printer away, of course, and displayed it like a trophy for the newsies. Its little shrine in the kitchenette seemed horribly empty. When I roused myself and picked up the

flat and rescued my poor peeping tweetybird, I put a blender there. It was made out of printed parts, so it would only last a month before I'd need to print new bearings and other moving parts. Back then, I could take apart and reassemble anything that could be printed.

By the time I turned 18, they were ready to let Da out of prison. I'd visited him three times — on my tenth birthday, on his fiftieth, and when Ma died. It had been two years since I'd last seen him and he was in bad shape. A prison fight had left him with a limp, and he looked over his shoulder so often it was like he had a tic. I was embarrassed when the minicab dropped us off in front of the estate, and tried to keep my distance from this ruined, limping skeleton as we went inside and up the stairs.

"Lanie," he said, as he sat me down. "You're a smart girl, I know that. You wouldn't know where your old Da could get a printer and some goop?"

I squeezed my hands into fists so tight my fingernails cut into my palms. I closed my eyes. "You've been in prison for ten years, Da. Ten. Years. You're going to risk another ten years to print out more blenders and pharma, more laptops and designer hats?"

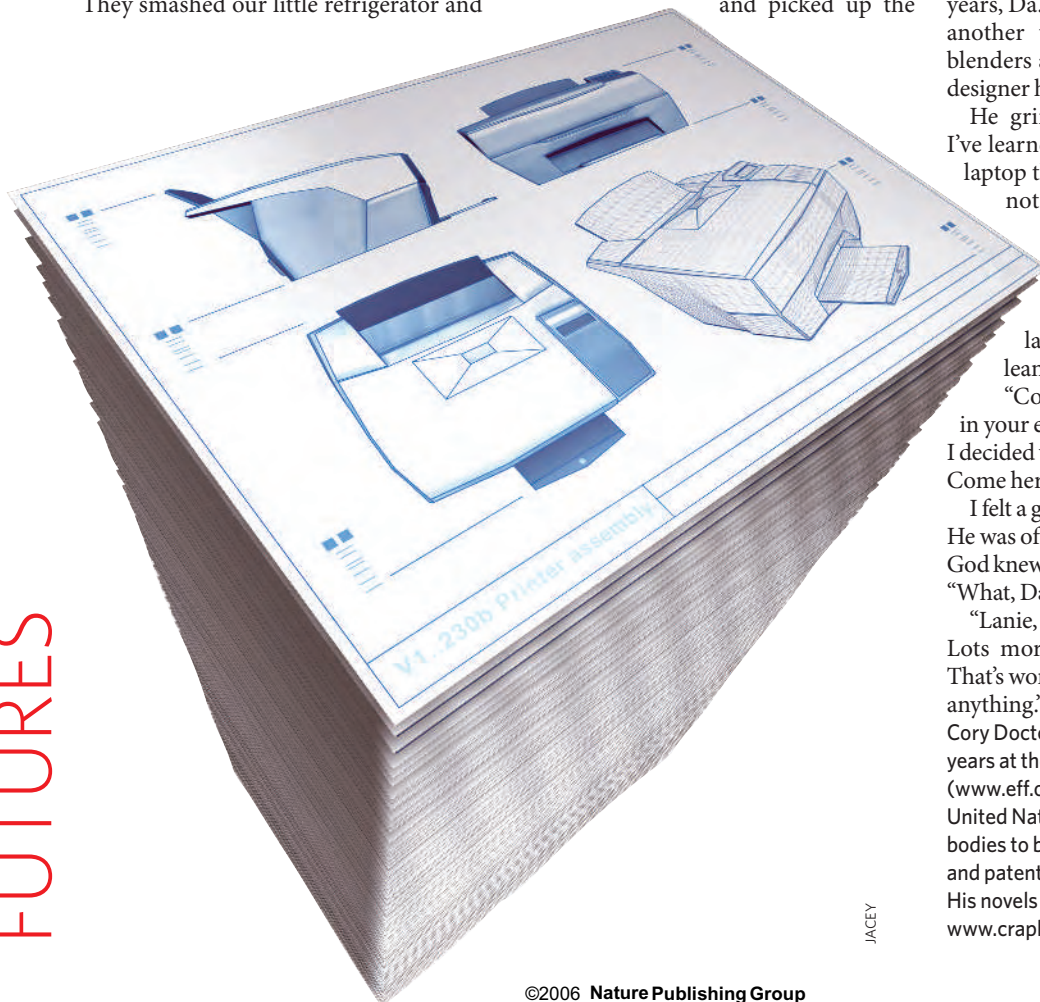
He grinned. "I'm not stupid, Lanie. I've learned my lesson. There's no hat or laptop that's worth going to jail for. I'm not going to print none of that rubbish, never again." He had a cup of tea, and he drank it now like it was whisky, a sip and then a long, satisfied exhalation. He closed his eyes and leaned back in his chair.

"Come here, Lanie, let me whisper in your ear. Let me tell you the thing that I decided while I spent ten years in lockup. Come here and listen to your stupid Da."

I felt a guilty pang about ticking him off. He was off his rocker, that much was clear. God knew what he went through in prison. "What, Da?" I said, leaning in close.

"Lanie, I'm going to print more printers. Lots more printers. One for everyone. That's worth going to jail for. That's worth anything."

Cory Doctorow has spent the past four years at the Electronic Frontier Foundation (www.eff.org), fighting at the United Nations and in tech-standards bodies to balance the rights of copyright and patent holders with the public interest. His novels can be had free online at www.craphound.com.



JACEY